The Causes of Ukrainian Famine Mortality, 1932-33*

Andrei Markevich†, Natalya Naumenko‡ and Nancy Qian§

February 22, 2024

We construct a novel panel dataset for interwar Soviet Union to study the causes of Ukrainian famine mortality (*Holodomor*) during 1932-33 and document several facts: i) Ukraine produced enough food in 1932 to avoid famine in Ukraine; ii) mortality in 1933 was increasing in the pre-famine ethnic Ukrainian population share and iii) was unrelated to food productivity across regions; iv) this pattern exists across the Soviet Union, even outside of Ukraine; v) migration restrictions exacerbated mortality; vi) actual and planned grain procurement were increasing and actual and planned grain retention (production minus procurement) were decreasing in the ethnic Ukrainian population share across regions. The results imply that anti-Ukrainian bias in Soviet policy contributed to high Ukrainian famine mortality, and that this bias systematically targeted ethnic Ukrainians across the Soviet Union.

JEL: N4, P2

Keywords: Ethnic Repression, Mass Killing, Famine

*We are grateful to the detailed comments of the editor, Elias Papaioannou, and the anonymous referees. We thank Daron Acemoglu, Maxim Ananyev, Volha Charnysh, Michael Ellman, Ruben Enikolopov, Jeff Frieden, Scott Gehlbach, Mikhail Golosov, Sergei Guriev, Mark Harrison, Oleg Khlevniuk, Alexey Makarin, Victor Malein, Joel Mokyr, Cormac O'Grada, Elias Papaioannou, Steven Nafziger, Maria Petrova, Sakari Saaritsa, Torsten Santavirta, Marco Tabellini, Chris Udry and Ekaterina Zhuravskaya for their insights; the participants of the Harvard/MIT Development Workshop, Harvard Political Economy Workshop, Duke University Development Workshop, NYU Development Workshop, LSE STICERD Applied Workshop, Northwestern Economic History Workshop, USC Development Workshop, Fudan University SOE Applied Workshop, PKU NSD Development Workshop, University of Pittsburg Development and Labor Workshop, Trinity College Applied Economics Workshop, CERGE-EI Applied Economics Workshop, EIEF Workshop, HECER (Helsinki Center of Economic Research) Labor and Public Economics seminar, Helsinki Economic and Social History Seminar, Humboldt University Economic History Seminar, Northwestern Economic History Lunch, Northwestern Development Lunch, George Mason Economic History Lunch, the NBER Political Economy Workshop, the Ridge Conference, the ASSA/AEA On the Role of and Culture and Community 2020 session, 2020 Summer Workshop in the Economic History and Historical Political Economy of Russia, Kiel/CEPR Conference on Geopolitics. For excellent research assistance, we thank Pavel Bacherikov, Nikita Karpov, Ivan Korolev, Sergio Lopez-Araiza Bejar, Kaman Lyu, Carlo Medici, Ludovica Mosillo. All mistakes are our own.

†University of Helsinki and New Economic School (Moscow, Russia), andrei.markevich@helsinki.fi, amarkevich@nes.ru, CEPR

‡George Mason University, nnaumenk@gmu.edu

§Kellogg School of Management at Northwestern University, nancy.qian@kellogg.northwestern.edu, NBER, BREAD, CEPR

Lazar Kaganovich "was fond of saying that every Ukrainian is potentially a nationalist" – from the memoirs of Nikita Khrushchev (Khrushchev, 1970, p. 172).

1 Introduction

During the Great Soviet Famine (1932-33), approximately seven million people perished and approximately forty percent of these deaths occurred in Ukraine, where mortality rates were four to six times higher than in Russia. 7.5% to 11.3% of the ethnic Ukrainian peasantry died during the *Holodomor* ("to die by starvation" in Ukrainian).[1] Levchuk et al. (2020) summarizes the controversy and intense debate about the causes of disproportionately high Ukrainian mortality rates. On the one hand, many historians believe that the famine was a "terror" intentionally waged by the Soviet government against Ukraine (e.g., Conquest, 1986). Approximately 40% of Soviet GDP was agriculture in 1928. To maximize revenues, the government forcibly procured agricultural production, the most important of which was grain, from rural areas. This was highly unpopular amongst the peasantry. The Ukrainians, who constituted the largest ethnic group in agriculturally productive regions, had a strong group identity and offered particularly stiff resistance. The Bolsheviks needed to repress Ukrainians to control agriculture (e.g., Conquest, 1986; Graziosi, 2015).

On the other hand are those who argue that the high famine mortality in Ukraine was an unintended consequence of policies with no inherent anti-Ukrainian bias. No direct documentary evidence that Stalin "ordered" a famine for Ukraine has been uncovered.[2] Repression and starvation in agriculturally productive regions also occurred outside of Ukraine. Communist ideology, in an intentional departure from the Tsarist regime, held all ethnicities to be equal. In their seminal economic history study of the Soviet Famine, Davies and Wheatcroft (2009) document a fall in aggregate Soviet production in 1932. They argue that mortality was higher in Ukraine because it was agriculturally productive and suffered bad weather and the Soviets had no anti-Ukrainian bias *per se*.

The contentious debate reached an *impasse* because the lack of representative disaggregated data prevented a systematic evaluation of competing hypotheses. Most importantly, past studies have lacked systematic disaggregated data on mortality, production and procurement.

The primary contribution of our study is to make progress on understanding the causes of Ukrainian famine mortality in 1932-33 by constructing the largest and most comprehensive disaggregated dataset for inter-World War Soviet Union (1922-1940). Most of the data are manually digitized from archival sources made available after the fall of the USSR. The main sample includes the three largest and most populous Soviet republics: Russia, Ukraine and Belarus. We construct a province-level panel data that includes information about mortality, natality, ethnic composition, urbanization, weather, administrative capacity, realized grain production and procurement, planned targets for production and procurement, political alignment with

[1] See Section 2 for mortality estimates. Note that approximately 1 to 1.5 million famine deaths occurred in Kazakhstan, mostly among ethnic Kazakhs. We do not study Kazakh mortality because there are no reliable mortality data from Kazakhstan during the famine era.

[2] Kondrashin (2008) points out that famine was severe in certain parts of Russia. Kotkin (2017) makes no mention of bias against Ukrainians. He argues that the Soviet famine was an unintended consequence, and notes the contrast between the lack of direct evidence on Stalin's intention to produce a famine for Ukrainians and the abundance of direct evidence for other ethnic minorities killings during the Great Purge. Tauger (1991) argues that there was no policy against Ukrainians or other ethnic groups.

Bolsheviks, and other historical economic, political and cultural variables. We use two approaches to reconstruct grain production data, which were manipulated by the Soviet government in the late 1920s and early 1930s*: i*) from previously classified grain procurement ratios, and *ii*) by predicting production with data that are hard to manipulate (e.g., weather and geography). We also construct a more granular district-level panel that contain data on mortality, demographic characteristics, weather and geography. The main advantage of these data is that they allow us to evaluate competing explanations for the famine and provide direct positive evidence on policy mechanisms.

Our study proceeds as follows. First, we document that 1932 Ukrainian production, though lower than previous years, was still enough to support all of Ukraine. Government procurement lowered food availability in rural Ukraine to a level that was similar to the famine-stricken regions of the 1921 famine. Procurement from Ukraine was not necessary for avoiding famine in other regions because production in the other republics was enough to support the rest of the Soviet population.

Second, we investigate the contributions of anti-Ukrainian bias in Soviet policy versus a fall in agricultural production caused by weather. These two causes of famine are not mutually exclusive. There is no data on ethnic-specific mortality rates. To overcome this challenge, we infer ethnic Ukrainian mortality from the relationship between famine mortality and the pre-famine ethnic Ukrainian population share across regions. Using the province-level panel, we document that for two places that produced the same amount of grain in 1932, the place with more ethnic Ukrainians suffered higher mortality in 1933. This supports the anti-Ukrainian bias view. At the same time, we find that for two regions with the same ethnic Ukrainian population share, per capita grain production is unrelated to famine mortality. Thus, grain production and by logical extension, the inputs of grain production (e.g., weather and geography), cannot explain famine mortality. This contradicts the unintended consequence view.

Our estimates account for urban population share and its interaction with the famine dummy variable, which account for urban-rural differences such as food access; province fixed effects, which account for all time-invariant differences across regions (e.g., average mortality rate); and year fixed effects, which account for changes over time that affect all provinces similarly (e.g., macroeconomic changes). We show that the results are robust to a large number of additional controls, different ways of measuring famine severity and Ukrainian population, and placebo experiments. Moreover, we document that Ukrainian population share is positively associated with famine mortality across districts within provinces as well as across provinces. This supports the interpretation that Ukrainian bias in famine mortality is policy driven since central government policies are implemented top-down and usually follow similar guidelines at each level of government.

A back-of-the-envelope calculation shows that if ethnic Ukrainians died at the same rate as other ethnicities (mostly Russians in our sample), total famine mortality would have been 92% lower in Ukraine and 77% lower in the three republics in our sample. The contribution of Ukrainian bias to famine mortality is higher in Ukraine because Ukraine has a higher share of ethnic Ukrainians.

Third, Ukrainian bias was specific to ethnic Ukrainians and present outside of Ukraine. We find a similar positive relationship between pre-famine Ukrainian population share and famine mortality in Belarus and Russia, where 25% of the ethnic Ukrainians in our sample resided prior to the famine. Other ethnic minorities, on average, had famine mortality rates similar to ethnic Russians, and much lower than Ukrainian

famine mortality rates.

Fourth, we use the district-level data to document a discrete decline in famine mortality rates when crossing the border from Ukraine to Russia. This is consistent with the belief that migration restrictions exacerbated famine mortality by preventing Ukrainians from escaping to other regions. If ethnic Ukrainians were allowed to freely migrate, mortality should not have discretely declined at the border. We then show that the border effect disappears once we control for the ethnic Ukrainian population share in each district. Thus, Ukrainian bias in famine mortality was delineated along ethnic lines.

Fifth, we provide positive evidence that anti-Ukrainian bias in famine mortality is driven by policy. We show that there was bias in the most important policy determining food availability: centrally planned grain procurement. We document that for two provinces that produced similar amounts of grain in 1932, the government procured more from the one with a higher share of ethnic Ukrainians. Thus, retention (production minus procurement) was decreasing in pre-famine ethnic Ukrainian population share. This result addresses the lack of documentary evidence for central directives ordering the famine. A simple quantification exercise suggests that anti-Ukrainian bias in centrally planned procurement explains approximately half of the total effect of anti-Ukrainian bias on famine mortality.

In addition, several supplementary results enrich our understanding of the famine. First, we document that there was anti-Ukrainian bias in the grain procurement targets published in 1928, which supports the presence of anti-Ukrainian bias in Soviet policy. Second, we document that after the famine, production recovered much sooner than population. The data on centrally allocated tractors suggest that the mechanization of agricultural was intensified to compensate for the negative impact of labor losses on production. Finally, we speculate about the regime's motivation and present some evidence consistent with the political economic explanation that the Bolsheviks repressed Ukrainians to control agriculture.

Related Literature Our study is the first to systematically evaluate the causes of Ukrainian famine mortality, to show that Soviet policy was biased against *ethnic* Ukrainians, and that this bias extended to republics other than Ukraine. These results add to the large literature on the causes of famine in the post-Industrial era discussed by Ó Gráda (2009). Sen (1981) famously argues that the central cause of 20th century famines is the unequal distribution of food from political elites to those who lack entitlement, and not low aggregate production. Our results strongly support this thesis by providing evidence on the detailed process that killed more than any other famine in the 20th century (except the Great Chinese famine of 1959-1961). Studying the Soviet context adds important insights. In market economies, famine mortality is negatively associated with food production (Sen, 1981). Our finding of no correlation between food production and famine mortality illustrates that different mechanisms are at play in the Soviet context. The lack of a negative association between food productivity and famine intensity is also notably missing in the Great Chinese Famine (Meng et al., 2015). Two key differences between the two famines of centrally planned economies is that in China, the rural political bases for the political leadership suffered high famine mortality rates and there was little ethnic delineation in mortality.

We complement existing studies of the Soviet Famine. The most well-known study in economics is Davies and Wheatcroft (2009), which pieces together the aggregate data with documentary evidence and

select variables from select regions. By conducting a systematic analysis, we arrive at very different conclusions from this seminal study about the causes of high Ukrainian mortality. More recently, Naumenko (2021) uses one cross section of districts in Ukraine and finds a positive association between famine mortality and ethnic Ukrainians share. Her study cannot distinguish normal mortality from excess mortality due to famine, lacks variation to meaningfully control for alternative hypotheses and data on production and procurement to examine policy mechanisms.

Our results also add to the mostly theoretical literature on authoritarian governance that is reviewed by Egorov and Sonin (2020) and the rapidly growing evidence on the economic determinants of mass killings and genocide. Our findings support the theory proposed by Esteban et al. (2015) that mass killings are more likely in the presence of large natural resource rents, political polarization, institutional constraints regarding rent sharing and low productivity of labor. They complement recent empirical evidence on the drivers of mass killing in contexts such as Rwanda (Heldring, 2020; Rogall, 2021; Yanagizawa-Drott, 2014) and Gujarat (Jha, 2014). Our results are consistent with the argument that persecutions are often triggered by negative economic shocks in places with some pre-existing prejudice, and can transform into genocide if the repressed group has no exit option (Becker et al., 2022). Finally, we add to studies of the long-run impact of early Soviet economic policy. Cheremukhin et al. (2017) provides macro-calibration evidence on the effect of early Soviet Industrial policy, but excludes the cost of famine because of data limitations. See Zhuravskaya et al. (forthcoming) for a review. We hope that the data we constructed will facilitate future studies about this important context.

This paper is organized as follows. Section 2 summarizes the historical background. Section 3 presents the food accounting exercise. Section 4 presents mortality estimates. Section 5 presents evidence of the role of government policy. Section 6 presents the long-run trends for rural population and agricultural production after the famine. Section 7 concludes.

2 Background

2.1 Ethnicity in the Soviet Union

In a departure from explicit ethnic discrimination during the Tsarist regime, Bolshevik ideology held all ethnicities to be equal. Yet, from the beginning of Soviet rule in 1917, the Bolsheviks needed to balance their reliance on the cooperation of nationalist groups with the concern that nationalist sentiments would undermine the regime. This was especially true for Ukrainians, who were the second largest ethnic group after Russians in the Soviet Union and the largest ethnic group on agriculturally productive lands.

The Ukrainians posed a dilemma for the Bolsheviks. On the one hand, the Bolsheviks needed their cooperation to rule. In the only free election in the Soviet era in 1917, the Party of Socialist Revolutionaries obtained the largest vote share of 39.2% of the former Russian Empire. The Bolsheviks took second place with 23.8% of votes, while the Ukrainian Socialist-Revolutionary Party took third place with 12.4% of the votes (estimated from Protasov et al. 2014). The Bolsheviks also needed Ukrainian cooperation to achieve its economic goal of rapid industrialization, which required subsidies from agriculture, the most important source of Soviet economic production at the time. The most important agricultural commodity at the time

was grain. On the other hand, the Bolsheviks were aware of the danger of nationalist sentiments to the regime and the fact that the Soviet rural economic policy of maximizing grain extraction from rural areas would be highly unpopular.

The strength of nationalist sentiments was apparent in the Civil War of 1918-1922. To appease nationalists, in 1923, the Bolsheviks launched a policy of *indigenization* (*korenizatsiya*). Indigenization encouraged schools and books in local languages, promoted native culture (e.g., national literature, theaters, museums), required local government affairs to be implemented in the native language, and promoted locals into leadership positions. The Ukrainian communist party was charged with the administration of Ukraine, but also viewed itself as representing ethnic Ukrainians across the Soviet Union.

The salience of ethnicity increased over time in rural areas. In the Russian empire, rural ethnic groups lived in separate communities. Ethnic delineations were inadvertently deepened during the Soviet regime, which established a hierarchy of national autonomous administrative units (republics, provinces, districts and villages) delineated along ethnic lines. The residential patterns and administrative structure encouraged organization and coordination along ethnic lines, and also made it logistically easier for government to implement ethnic-specific policies.

The Bolsheviks leaders were wary of the increasing salience of ethnicity in rural areas and their concerns increased when peasants began to resist agricultural collectivization in the late 1920s and early 1930s. Resistance was particularly strong amongst Ukrainians. Indigenization de facto ended in Ukraine and other European parts of the USSR in the autumn of 1932 (Graziosi, 2015; Martin, 2001).

2.2 Soviet Economic Policy and the Famine

Rapid industrialization funded by agricultural production was the cornerstone of Soviet economic policy in the late 1920s and 1930s. The government aimed to maximize the expropriation of grain surpluses from the countryside. The first attempt took place amidst the Civil War (1918-1922), which was fought mainly between the Bolshevik-led red army, the white army and separatists. *War communism* limited trading of foodstuffs and introduced *prodrazverstka* aimed to extract all 'surplus' grain from peasants. The peasants resisted by decreasing sown area, which in 1921 was 30% lower than in 1913. The disruptions from War Communism and the armed conflict contributed to the 1921 famine, in which approximately five million died. Most deaths were in the Volga region in Russia (Andreev et al., 1993). The scale of the resistance and the decline in agricultural production, which constituted over half of Soviet GDP in the early 1920s threatened to undermine the Bolshevik economy. In 1921, Lenin declared the New Economic Policy, which re-introduced a market economy for agriculture and small-scale manufacturing. The failure of *War communism* made clear to the Bolsheviks that they needed more control over agriculture.

In 1928, the Soviet Union was led by Stalin, who substantially consolidated political power. The Bolsheviks, no longer distracted by armed conflict, renewed their effort to control agriculture. *Collectivization* was a bundle of policies that included the removal of individual family farms and organizing peasants into large collective farms. The government started limiting and heavily regulating the trading of food and directly procuring it from peasants to feed the urban population and the small share of rural population engaged in non-agricultural production (e.g., forestry). Procured foods, mostly grain, were also exported and stored

in centrally controlled reserves. Production and procurement targets were set by central planners and published in 1928 in the First Five Year Plan. Production targets were mainly based on production potential, which depended on past production, geography and rural labor. Procurement targets were set to leave the rural population enough food to be productive laborers. Forced collectivization began in late 1929. By the summer of 1932, the share of rural households living in collective farms exceeded 60% in the USSR and was almost 70% in Ukraine (Davies and Wheatcroft, 2009).

The harvest of 1931 was lower than earlier years. News of starvation traveled to Moscow, but the government did not lower procurement. In areas where the peasants retained too little food, seed stock was consumed to make up for the deficit. This, in turn, contributed to lower production in 1932. It is widely believed that the production decline was particularly prominent in Ukraine, but the exaggerated official production data during the early 1930s made it hard to verify the magnitude of the fall in each region.

Facing lower than expected harvests in 1932, Stalin simultaneously curtailed the initial targets and emphasized the need to maximize procurement. To fulfill the remaining quota, Stalin sent his closest deputies, Vyacheslav Molotov and Lazar Kaganovich (neither of whom were ethnically Ukrainian) to Ukraine and North Caucasus, the two key grain-producing regions where most ethnic Ukrainians lived.[3] On December 14, 1932, the Politburo issued a classified decree to accuse Ukrainian nationalists within the Communist Party and local bureaucracy of sabotaging grain procurement. The decree required regional authorities in Ukraine (as well as in the North Caucasus and the Western region) to "crush" any resistance of "counter-revolutionaries" and nationalists and fulfill procurement quotas (Danilov et al., eds, 1999-2006, Volume 3, Document 226).

Deaths from starvation peaked in the early months of 1933. In January 1933, Moscow ordered the closure of the borders of Ukraine and the North Caucasus to prevent a mass migration of peasants out of these areas (Danilov et al., eds, 1999-2006, vol.3, p.634-5). The government gave little aid. Ukraine received 1.3% of what was taken in 1932, and it was mostly in the form of seeds for production in the following year rather than food (Davies and Wheatcroft, 2009, Table 23). Some cities also provided other forms of famine relief, such as aid kitchens, medical assistance or housing on an ad hoc basic. According to anecdotal accounts, rural famine victims often went to nearby urban areas to beg for food; some set up relief kitchens, while others expelled migrants back to rural areas. There are no systematic data about such relief efforts.

Total famine mortality estimates for the Soviet Union range from 5 to 10.8 million. Mortality was concentrated in rural areas.[4] National mortality rates returned to trend in 1934.

[3] In a letter to his deputy, Lazar Kaganovich, from August 11, 1932, Stalin mentioned that the party district committees in about fifty districts in Ukraine had spoken out against state procurement quotas and that the Soviet government "could lose Ukraine" (Davies et al., eds, 2003). The central leadership claimed that shortages and famine were outcomes of intentional peasant resistance aimed to undermine agricultural collectivization, and that the peasants should be penalized for their subversion (Danilov et al., eds, 1999-2006; Davies and Wheatcroft, 2009). On May, 6 1933, during the peak famine mortality, Stalin wrote to Sholokhov, a famous writer originally from the Don region, that peasants "sabotaged" his policy and accused them of engaging in a "silent war" against the Soviet state (Murin, ed, 1997). Villages were penalized for failing to fulfill procurement targets with the seizure of other foodstuffs (not just grain), bans of imports of foodstuffs and manufacturing goods to these villages, arrests of local government bureaucrats, and the deportations of peasants (Zelenin et al., 1994, p. 258, 260).

[4] Conquest (1986) estimates total famine deaths to be 7 million. Davies and Wheatcroft (2009) estimates 5.5 to 6.5 million deaths. Ellman (2005) cites "'about eight and a half million' victims of famine and repression in 1930–33." Kondrashin (2008) gives a range between 5 and 7 million victims. Russian historical demographers estimate 7.2 to 10.8 million famine victims (Polyakov and

There are no systematic data on ethnic-specific mortality rates in the Soviet Union for the period of our study. One way to approximate ethnic Ukrainian famine mortality is to use the most cited total famine death toll of seven million for the USSR (Conquest, 1986), and 2.6 million (Meslé et al., 2013) to 3.9 million (Rudnytskyi et al., 2015) for Ukraine. If famine deaths were equally distributed between ethnic Ukrainians (80% of Ukraine) and others ethnicities in Ukraine, and no ethnic Ukrainians died outside Ukraine, then ethnic Ukrainian deaths constitute 30% ($.8 \times 2.6/7 = .3$) to 45% ($.8 \times 3.9/7 = .45$) of the total famine deaths. Thus, ethnic Ukrainians, who were 21% of total Soviet population in 1926 constituted about 40% of all famine deaths. These estimates likely underestimate ethnic Ukrainian deaths because ethnic Ukrainian mortality was higher than those for other groups within Ukraine and many Ukrainians who lived in other republics also died.

Another way to assess Ukrainian famine mortality is to compare famine mortality rates in Ukraine to those in Russia. If total famine deaths is seven million and we subtract the deaths in Kazakhstan (1 to 1.5 million) and Ukraine (2.6 to 3.9 million), we are left with 1.6 to 3.4 million deaths for Russia if we assume no famine mortality in other republics. This implies famine mortality rates of 14 to 30 per 1,000 for the 112 million residents of Russia. A similar calculation for Ukraine yields a famine mortality rate that is four to six times higher than in Russia: 81 to 122 per 1,000.

In summary, any calculation indicates that Ukrainians suffered higher mortality rates during the famine than Russians. But we do not know if this is simply because Ukrainians lived more in agricultural regions which suffered a fall in production. We do not know if Ukrainians simply produced too little food in 1932 and would have suffered severe famine in 1933 even absent government procurement. The historical evidence indicates that the Bolsheviks were concerned about Ukrainian nationalism, but does not reveal whether they intended to procure more food from Ukrainians above and beyond other peasants in similarly productive lands. The historical narratives focus on the tension between Moscow and Ukraine, but say little about ethnic Ukrainians living elsewhere in the Soviet Union. The subsequent empirical analysis will address these and other issues.

3 Food Accounting

This section conducts a simple republic-level accounting exercise and documents that had there been no procurement from Ukraine, 1932 grain production in Ukraine was sufficient to avoid famine in Ukraine and production in the rest of the Soviet Union was sufficient to avoid famine there without grain from Ukraine. Table 1 presents grain production, procurement, retention (production minus procurement) and compares retention to rural population food needs for years from 1927 to 1939. These data cover the entire Soviet Union. Panel I examines Ukraine. Rows (1) and (2) report total and rural populations measured at the beginning of each year. The 1933 population data are taken before the famine mortality peaks.

Zhiromskaya, 2000). In 2008, the Russian State Duma postulated that within the territories of the Volga Region, the Central Black Earth Region, Caucasus, Ural, Crimea, Western Siberia, Kazakhstan, Ukraine and Belarus, the estimated famine death toll was 7 million people (State Duma, 2008). The differences in estimates are driven by data limitations, and potential underregistration of deaths during the famine. Estimates deriving excess famine deaths from a comparison of the pre-1926 and post-1937 Soviet population censuses suffer from the problem of the underregistration of infant mortality (See Davies and Wheatcroft, 2009, for a detailed discussion).

Row (3) reports production. The main challenge for this exercise is that the aggregate production data are widely believed to have been exaggerated during the early years of collectivization because they were publicly used as a marker of the success of Soviet economic policies. Davies and Wheatcroft (2009) provides a range for true aggregate production for the Soviet Union. We follow the spirit of this earlier study and use previously classified data to construct corrected province-level production, which we can then aggregate to the republic and the whole Soviet Union levels. Specifically, we use grain procurement ratios (grain procurement as a share of production) and procurement stocks to back out true production. Procurement stocks were directly observed and counted by the government. Procurement ratios were reported in Gosplan, a candid evaluation of the First Five-Year Plan shown only to the highest ranking Soviet officials. The report was classified until after the fall of the Soviet Union and was only recently discovered by historians. Our correction assumes that changes in the procurement ratio reflects changes in production rather than changes in the ability to procure. The corrected production measures are lower than officially reported production during the years when production is believed to have been exaggerated. At the Soviet Union level, our measures are comparable to the production estimates provided by Davies and Wheatcroft (2009). See Appendix Section A for details.

Row (4) presents grain procurement, which includes urban consumption, exports and national reserves. Production and procurement are reported in millions of tons.

Row (5) reports actual rural retention, the difference between production and procurement, which includes seed stock intended for cultivation (since peasants consume seed stock during times of starvation). We report retention in kilograms per capita per day which we find to be more intuitive for considering subsistence needs. Row (6) reports a counterfactual retention for rural Ukrainians which assumes that no grain is procured from rural Ukraine. Row (7) reports a counterfactual retention for urban and rural Ukrainians which assumes that grain produced in Ukraine is distributed only to those living in Ukraine. This is simply production divided by the total population, converted into kilograms per capita per day units.

To understand whether retention is sufficient to avoid famine, we calculate population food needs (row 8). We conservatively use official Soviet guidelines for maximum caloric needs that vary by sex, age and occupation (Lositskij 1926; 1928). We use data on sex, age and urban-rural population shares from the 1926 Census to adjust subsistence needs according to the demographic composition of each republic.[5]

The data show that grain production in Ukraine declined from 23.2 to 16.8 million tons from 1930 to 1931, and further declined to 9.1 million tons in 1932 (row 3). Grain procurement from Ukraine remained stable at 7.7 and 7.3 million tons in 1930 and 1931; and decreased to 4.2 million tons in 1932 (row 4). The decline in procurement is consistent with the earlier historical discussion that the regime was unable to procure initial targets in the face of the production drop in 1932. Grain procurement declined with the drop in production in 1932, but not enough to prevent famine. Per capita grain retention in rural Ukraine declined from 1 kg per capita per day after the harvest of 1931 to only 0.5 kg per capita per day after the harvest of 1932 (row 5). This was 30% below the estimated food needs of 0.778 kg per capita per day (row 8). For another comparison, note that during the 1921-22 famine, famine occurred in regions with per capita

[5]The Soviet guidelines are more generous than international standards. For example, for a prime-age adult male, it is assumed that 3,750 calories are needed for heavy labor. We convert calories to kilograms of grain using the conversion offered by Lositskiy (1920), which is based on the typical grain consumption of Russian laborers.

retention below 0.628 kg per person per day (Bukhman, 1923).

The famine in rural Ukraine could have been avoided, or at least, greatly moderated if there was no procurement from rural Ukraine. If no food had been procured from rural Ukraine, rural retention would have been 1.01 kg per capita per day in 1932 (row 6), much more than the recommended level of 0.778 kg per capita per day. Moreover, Ukraine produced enough food to feed all its citizens. If grain was procured from rural Ukraine only to feed urban citizens of Ukraine, average Ukrainian retention would have been 0.79 kg per capita per day (row 7).

Procurement from Ukraine was not necessary to avoid famine elsewhere. Panel II examines all other republics of the USSR. Row (11) shows that urban and rural grain retention would have been 1 kg per capita per day in 1932 if the production of other republics was distributed equally across their population. Appendix Table A.2 also presents the food accounting for Russia and the entire Soviet Union.

There are several important points to keep in mind for interpreting the accounting results. First, we focus on grain because it is the most important agricultural commodity for the Soviet economy and the main staple for consumption. 66% of peasant calories came from grain, 15% came from potatoes, 9% from diary, and all other foods contributed only 10% together (Naumenko, forthcoming). We do not have disaggregated data for the production of other crops. Later, in the regression estimates, we account for this by controlling for the suitability for cultivating other staple crops, such as potatoes, as well as cash crops. The availability of other food would strengthen our point that there was enough food in aggregate to avoid famine. Second, our estimates do not take into account losses during transportation or storage because of data limitations. Lositskiy (1920) estimates such losses to be around 5%. Our threshold for food needs is very generous since the amount for labor is much more than the amount needed to avoid mortality. Thus, taking food losses into account should not overturn the result that famine could have been largely avoided if no food were taken out of Ukraine. Third, we do not account for food that the central government distributes back to regions after procurement. These data are not systematically available and they are unlikely to affect our results because the transfers were mostly sent to urban areas and rural workers not involved in grain production (e.g., forestry workers). The accounting exercise also excludes *ad hoc* famine relief efforts that we discussed in the Background Section. The small quantities of such relief mean that they are unlikely to overturn our result. Finally, the estimates in this section do not take into account the fact that some of the population living in Ukraine are not ethnically Ukrainian, or that many ethnic Ukrainians lived outside Ukraine. We address this in the next section by inferring ethnic-specific mortality rates from the regression estimates.

4 The Causes of Ukrainian Famine Mortality

4.1 Ukrainian Bias, Production Fall and Other Factors

Motivated by the current debate on the causes of famine in Ukraine, we examine the contributions of Ukrainian bias and regional food production.

Our main measure of famine severity is the (excess) mortality rate. Province-level mortality data are available for each year in our sample for nineteen provinces from the three most populous Soviet republics:

Belarus, Russia and Ukraine. These three republics include 84% of the 1926 Soviet population and 88% of the 1928 Soviet grain production. For brevity, we refer to these three republics as the Soviet Union or the USSR in the remainder of the paper. The average province in our sample has 6.5 million people in 1926. All data are mapped to their 1932 province borders.

Figure 1a plots mortality from 1923 to 1940 for Ukraine and the other two republics. It shows that during non-famine years, mortality rates are lower in Ukraine (18 per 1,000) than elsewhere (22 per 1,000) and stable over time. However, mortality sharply increases during the famine. In Belarus and Russia, mortality increases in 1933 to approximately 30 per 1,000. In Ukraine, the increase begins slightly earlier in 1932 to approximately 22 per 1,000 and then spikes in 1933 to approximately 60 per 1,000. The earlier timing and larger magnitude of the total mortality increase in Ukraine are consistent with historical accounts that starvation began earlier in Ukraine and the greater intensity of the famine peak in 1933. There is significant variation in famine mortality across and within provinces (see Appendix Figure A.1a).

The data on ethnic Ukrainian population share is reported by the 1926 Population Census (the last census before the famine). In the 1926 Census, 23.2 million ethnic Ukrainians lived in Ukraine and an additional 7.9 million lived in Russia and Belarus. Ukrainians constituted 21% of total Soviet population and made up the second largest ethnic group. Russians, the largest group, constituted 53% of the population. 89% of ethnic Ukrainians lived in rural areas. In regions that produced large food surpluses that the government designated as "grain-surplus" areas, Ukrainians were the largest group (43.8%) and Russians were a close second (41.9%). Since most non-Ukrainians in our sample are ethnic Russians, we sometimes refer to the reference group as "Russians".[6] Ethnicity is self-reported in the census and these data are widely accepted to be accurate. There was little reason to intentionally misreport.[7] Later, we show that our results are similar when using alternative measures of ethnicity.

Next, we present bin scatter plots of mortality against the ethnic Ukrainian population share. The y-axis is the number of deaths per 1,000 people. The x-axis indicates which percentile a province is at in the distribution of 1926 Ukrainian population share. Figure 1b shows that in non-famine years, the relationship is moderately negative: provinces with more ethnic Ukrainians experience slightly lower mortality. Figure 1c shows that during the famine, the relationship is strongly positive: provinces with more ethnic Ukrainians experience higher mortality.

The following equation characterizes the relationship between famine mortality, pre-famine ethnic Ukrainian population share and grain production.

$$mortality_{i,t+1} = \alpha + \beta Ukrainians_i \times Famine_t + \gamma grain_{i,t} \times Famine_t + \Gamma X_{it} + \eta_i + \delta_t + \varepsilon_{it}. \qquad (1)$$

Mortality in province i during year $t+1$ is a function of: the interaction of the rural ethnic Ukrainian

[6] Appendix Figure A.1b maps the share of ethnic Ukrainians in the rural population for each province as reported in the 1926 Census. Grain-surplus regions are shaded in crosses. Appendix Table A.3 presents the correlates of 1926 Ukrainian population share and political, social and economic variables.

[7] The 1926 Population Census is commonly viewed as one of the highest quality Soviet censuses (Andreev et al., 1998). It is the last census before agricultural collectivization. After Ukrainians, the next largest ethnic group was an order of a magnitude smaller: Belorussians were 3.2% of total Soviet Union population and 3.5% of our sample. Appendix Table A.1 Panel A lists the three largest ethnic groups in the entire Soviet Union, Panel B lists the three largest ethnic groups in our sample, and Panel C lists the three largest ethnic groups in the grain-surplus provinces of our sample.

population share in province i in 1926, $Ukrainians_i$, and a dummy variable that equals one in the famine year, $Famine_t$; the interaction of per capita grain production, $grain_{i,t}$, and the famine dummy; and a vector of additional controls, X_{it}, which include the uninteracted grain production variable, urban population share and its interaction with the famine year dummy; province fixed effects η_i; and year fixed effects δ_t. Since $Ukrainians_i$ is a time-invariant measure, the uninteracted term is absorbed by the province fixed effects. We define the famine dummy, $Famine_t$, to equal one in 1932 because 1933 was the year with the highest mortality rates and we assume that grain production in year t is used to sustain the population in year $t+1$. Our results are qualitatively similar if the famine variable takes the value of one for 1931 and 1932. This is clear when we present the dynamic estimates later in the paper.

Given the nature of food production and famine, we allow the correlation of the standard errors to gradually decay across space to account for the movement of people, information and other factors across space. Later, when we use geographically smaller districts as units of observation, we estimate both standard errors that are spatially corrected and clustered at the district level. See Appendix Section C for alternative estimates.

β is the difference in the correlation of Ukrainian share and mortality between famine and non-famine years. If ethnic Ukrainians died at higher rates during the famine, then $\beta > 0$. γ captures the relationship between per capita grain production and famine mortality rates. If famine mortality was caused by low grain production, then $\gamma < 0$. We focus on the interaction coefficient, γ, because the relationship between production and mortality is likely to differ between famine and normal years, when production levels were higher. This is because the positive relationship between food consumption and survival exhibits strongly diminishing returns.

To account for urban-rural differences in food policies and other factors that affect famine mortality, X_{it} includes urban population share in province i during year t, and its interaction with $Famine_t$.

Table 2 presents the results. We measure grain production in several ways to address the concern that official figures from the early 1930s are exaggerated. In column (1), we use the corrected estimates based on previously classified documents as in the accounting exercise. However, the data we use for the correction are only available for a few years (1928 to 1933) and the correction is subject to assumptions discussed in Section 3. Alternatively, column (2) presents estimates that control for production predicted by weather and geography. These data are available for a longer time horizon and allows us to avoid the assumptions needed for the corrected production measure. Predicted grain can be interpreted as a parsimonious way of controlling for weather and geography. See Appendix Section B for details.

In both columns (1) and (2), the interaction coefficient of Ukrainian population share and the famine dummy variable is positive and statistically significant at the 1% level. For two provinces with the same per capita production in 1932 (and urbanization), the one with a higher share of ethnic Ukrainian population prior to the famine experienced higher famine mortality. This is consistent with anti-Ukrainian bias in famine-era policies contributing to famine mortality. In contrast, the interaction coefficient for grain and the famine dummy variable is positive. Because the positive interaction coefficient varies in precision and magnitude with the two measures of production (and the slightly different samples), we will cautiously interpret it as a null result.

These estimates imply that for two places with the same ethnic Ukrainian population share (and urbanization), famine mortality was uncorrelated to production in 1932. This is inconsistent with the drop in production (and therefore, the drivers of production, such as earlier policies or weather) contributing to famine mortality. To maximize sample size, we henceforth focus our discussion and the remaining mortality regressions in this section on the estimates using predicted grain.

The implied magnitude of Ukrainian bias in famine mortality is large. Taken literally, column (2) implies that for two provinces that experienced the same weather, and which have the same degree of urbanization, famine mortality rate was higher by 51 per 1,000 in a province with 100% Ukrainians than in a province with no Ukrainians. Mean mortality rates are 21 per 1,000 during non-famine years and 31 per 1,000 during the famine. Another way to assess the magnitude is to examine the standardized coefficient, which is presented in italics. During the famine, increasing Ukrainian population share by one standard deviation would result in a 0.826 standard deviation increase in mortality relative to normal years.

Predicted grain productivity is a noisy measure of actual grain productivity, which can bias the estimates in column (2). The similarity in our findings when using corrected grain productivity in column (1) implies that this is unlikely to be a major issue. Nevertheless, to be cautious, we consider the possibility of non-random measurement error. Conceptually, controlling for predicted grain production controls for production inputs such as weather and the geographic and climatic suitability for cultivating grain. However, if the relationship between these factors and production changes between the pre-Soviet era and the Soviet era, then the predicted estimates may be misleading. In our context, one may be concerned that earlier Soviet policies reduced the returns to production inputs so that predicted production overstates true production in Ukrainian regions.

We address this by controlling for the two policies that were most likely to have reduced agricultural productivity. In columns (3) and (4), we examine the sensitivity of the ethnic Ukrainian interaction coefficient to controlling for *dekulakization* and the loss in livestock that occurred just prior to the famine.

In the *dekulakization* campaign, approximately two million peasants (*kulaks*) were exiled to Siberia and other remote regions for actively resisting collectivization (Viola, 2007). *Kulaks* were often the relatively more productive peasants and their removal could have reduced the returns to the inputs we use to predict production. Between 1929 and 1932, the number of horses declined by 42% and cattle by 40% (Viola, 1996, p. 70). When peasants lost the property rights to their livestock, they responded by slaughtering, eating or simply neglecting the newly collectivized animals. Livestock was the main source of horsepower and manure was an important input for crop cultivation. The loss of livestock could have reduced productivity and undermined the traditional way to avoid famine when harvests are low (by slaughtering and eating the animals).

We control for the number of *kulak* households exiled from each region in 1930-31 divided by the 1930 population (column 3) and the drop in per capita livestock between 1929 and July, 1931 (column 4). Since these variables are time invariant, we control for their interactions with the famine indicator. The interactions of Ukrainian population share and the famine dummy variable are similar to the baseline. Thus, the baseline result is unlikely to be confounded by systematic mis-measurement of production.

In column (5), we address the concern that province fixed effects absorb meaningful variation in ethnic

Ukrainian share and famine mortality by including the uninteracted Ukrainian effect instead. The ethnic Ukrainian interaction coefficient in column (5) is nearly identical to the baseline in column (2). The uninteracted Ukrainian coefficient is -0.007 and statistically significant at the 1% level. This is similar to the mortality figures in showing that in non-famine years, ethnic Ukrainian population share is *negatively* associated with mortality. It is only during the famine that mortality is *positively* associated with ethnic Ukrainian population share. The sum of the interaction and uninteracted coefficients presented at the bottom of the table is positive and statistically significant at the 1% level.

To investigate the dynamic patterns of the mortality-Ukrainian gradient and provide evidence against spurious correlations, we replace the interactions with the famine dummy for 1932 with interactions with dummy variables for each year. The reference year is 1923. Figure 1d plots the interaction coefficients and their 95% confidence intervals and shows a sharp temporal pattern that goes against the concern that the baseline estimate is confounded by spurious correlations (Appendix Table A.6). The correlation between Ukrainian population share and mortality becomes positive in 1932 and peaks in 1933. There is no correlation in other years.

One way to quantify the total contribution of Ukrainian bias in famine mortality is to conduct a simple back-of-the-envelope calculation using the estimates in Table 2 column (2). The regression predicts that the number of deaths is on average 2.72 million in non-famine years and 4.97 million in 1933. The number of excess deaths due to the famine is the difference between mortality during famine and non-famine years: $4.97 - 2.72 = 2.26$ million (with rounding error). We obtain the counterfactual famine mortality in a world with no anti-Ukrainian bias by setting the interaction coefficient of ethnic Ukrainian population share and the famine dummy variable in equation (1) to zero. When we do this, predicted deaths in 1933 is 3.23 million. The number of famine deaths without bias against ethnic Ukrainians is the difference between this number and the number of deaths in non-famine years, 0.51 million ($3.23 - 2.72 = 0.51$ million). Thus, bias against Ukrainians accounts for 77% ($1 - 0.51/2.26 = 0.77$) of famine deaths in our sample. Since most non-Ukrainians in our sample are Russians, who suffered much lower famine mortality rates, our results imply that total famine mortality would have been 77% lower if ethnic Ukrainians died at similar rates as ethnic Russians. When we repeat the calculation for only Ukraine, we find that bias against ethnic Ukrainians accounts for 92% ($1 - 0.12/1.51 = 0.92$) of famine deaths in Ukraine. The larger magnitude is due to the fact that ethnic Ukrainian population share in Ukraine is higher than for the three republics combined.

Later, we will use the production and procurement evidence to show that Ukrainian bias in famine mortality is in large part due to anti-Ukrainian bias in policies that contributed to mortality. It is important to note that the main data used in our analysis were collected by the Soviet government and available to central planners. Our interpretation assumes that the estimate of Ukrainian bias is not confounded by spurious factors (e.g., variables that are correlated with pre-famine ethnic Ukrainian share and famine mortality, but unrelated to anti-Ukrainian bias in famine-era policy) and that anti-Ukrainian bias in policy led to famine mortality. We provide evidence for this later in the paper: we show that Ukrainian bias in famine mortality is unlikely to be confounded by omitted variables, and provide positive evidence that there is anti-Ukrainian bias in food procurement policy.

4.2 Ukrainian Bias Outside of Ukraine, Other Ethnic Minorities

In 1926, 25% of ethnic Ukrainians in our sample lived outside of Ukraine, which is one province in our sample. Column (6) of Table 2 examines the Ukrainian-mortality gradient in other provinces by omitting Ukraine. The Ukrainian interaction coefficient is positive and statistically significant. Thus, the bias against ethnic Ukrainians in famine mortality extends across the Soviet Union, beyond Ukraine.

In column (7), we add an interaction of the pre-famine share of non-Ukrainian ethnic minorities (e.g., Belorussians, Tatars, Mordvins, Chuvashs, Germans, Bashkirs, Jews, Poles and all other non-Ukrainians and non-Russians) with the famine dummy variable. If bias against Ukrainians was part of a larger policy of bias against all ethnic minorities, we should find that this interaction coefficient is also large, positive and statistically significant. Instead, we find that it is negative, negligible in size and only marginally significant. This interaction coefficient becomes insignificant in the district-level estimates presented later in the paper. Thus, we will interpret this as a statistical zero. These results only show bias in famine mortality against ethnic Ukrainians.

4.3 Robustness

1892 Famine The key concern for interpreting anti-Ukrainian bias in famine mortality is omitted variables. One may wonder if there are certain characteristics about ethnic Ukrainian society or the places they reside that make areas with large shares of Ukrainians particularly vulnerable to famine. For example, a recent study by Buggle and Durante (2021) argues that social capital can play an important role for surviving famines. Or, it may simply be that because Ukrainians are used to producing more food, they are less equipped to address harvest shortfalls. We investigate these possibilities by examining the 1892 famine, the last large famine in the Russian empire, with mortality from 1885 to 1913 shared with us by Charnysh (2022). Column (8) of Table 2 shows that famine mortality is not associated with ethnic Ukrainian population share. Thus, the Ukrainian-mortality gradient is specific to the Soviet famine of 1932-33 and unlikely to be explained by slow-moving features of Ukrainian culture or institutions.

Alternative Measures of Famine Severity Our analysis focuses on mortality as the main measure of famine severity, Here, we examine two alternative measures of famine severity. The first is natality. Live births should be decreasing in famine severity since starvation is negatively associated with the probability of a healthy pregnancy or birth, and is positively associated with the probability of miscarriage and stillbirths (Dyson and Ó Gráda, eds, 2002). The second is birth cohort size measured in the 1939 Census as the dependent variable. Following the method of Meng et al. (2015), we use the place of residence and age in 1939 to create a synthetic panel of province-specific birth cohort sizes. The birth cohort size sample has more observations because provinces in the 1939 Census are smaller than those we use in the earlier samples. We do not have annual population data for these smaller units and therefore normalize cohort size with 1939 total province population. We control for 1926 urban population share interacted with the famine instead of time varying urban population share and its interaction with the famine.

Columns (9) and (10) of Table 2 present the natality and birth cohort size estimates from the baseline specification. The interaction coefficients are all negative and statistically significant at the 1% level. The

estimates are consistent with the main finding of anti-Ukrainian bias in famine mortality. A back-of-the-envelope exercise using the natality and birth cohort size estimates show that anti-Ukrainian bias explains 54% of missing births and 50% of missing survivors in 1939 in Ukraine, and 26% of missing births and 14% of missing survivors in 1939 in Belarus, Russia and Ukraine.

In Appendix Section D, we present additional robustness checks, such as directly controlling for weather, using alternative measures of Ukrainian population share, controlling for the cultivation of other crops, demographic structure, administrative capacity, political zealousness and social norms. The robustness results taken together imply that our interpretation would only be confounded by a factor that is not accounted for by the large number of controls and which only matters in the Soviet regime (but not the Tsarist regime).

4.4 Within-Province Patterns

Soviet policies were centrally planned, implemented top-down by the bureaucracy and usually used similar allocation rules at various administrative levels. If the patterns we observe are driven by centrally planned policies, then we would expect to see positive famine-mortality-Ukrainian-share gradient across districts within provinces as well as across provinces. Our district-level panel consists of two years: 1928 and 1933. The majority of the data are manually collected from former Soviet archives. District-level mortality data are only available for Russia and Ukraine and we have fewer variables at this more disaggregated level (e.g., there are no data for production or procurement).

Table 3 column (1) re-estimates the baseline specification with district and year fixed effects instead of province and year fixed effects. Conceptually, famine mortality in this regression is the difference between 1933 and 1928 mortality rates. Since we do not observe production at the district level, the baseline controls for the suitability for grain cultivation from the FAO GAEZ database and its interaction with the famine year dummy; and weather (monthly temperature and precipitation in years t and $t-1$).

Column (2) and all subsequent columns control for province-year fixed effects, which isolate the within-province variation and control for factors that vary by province and year (e.g., regional political competition, leadership differences across provinces). The spatial patterns are similar to the province-level estimates. Within provinces, famine mortality is increasing with ethnic Ukrainian population share.

Column (3) controls for an alternative measure of weather: the deviations from long-term (1900-1950) means of monthly temperature and precipitation in years t and $t-1$. The results are similar. In column (4), we omit Ukraine and show that the patterns are similar in Russian provinces.

In column (5), we show that there is no bias for other ethnic minorities. The coefficient is small in magnitude and statistically insignificant.

The results are consistent with the presence of a systematic and centrally planned policy that targets ethnic Ukrainians.

The maps in Appendix Figures A.1c and A.1d illustrate the variation in ethnic Ukrainian share and famine mortality across districts within provinces. Appendix Table A.8 shows that the results are qualitatively similar when we subject the district-level estimates to the same sensitivity checks as the province-level estimates.

4.5 Administrative vs. Ethnic Boundaries

Past discussions about Ukrainian famine mortality have focused on the difference between Ukraine and other republics. Yet, our estimates show that Ukrainian bias in mortality exists outside of Ukraine. In this section, we use the district-level data to directly investigate the importance of administrative borders by examining the change in district-level famine mortality as one crosses from Ukraine to Russia. We plot famine mortality, the difference between 1933 and 1928 mortality rates, against the distance to the border between Ukraine and Russia, together with the fitted lines and their 95% confidence interval. Figure 2a shows that there is discrete decline in famine mortality rates as one crosses the border from Ukraine to Russia. Famine mortality rates are lower in Russia by 2.5 to 3.6 percentage-points (see Appendix Table A.9 Panel A).

The border effect on mortality rates is consistent with survivor accounts of notably lower mortality across the border (e.g., Applebaum, 2017, Ch. 10, 11). It also sheds light on the contribution of the migration ban on Ukraine imposed in January, 1933. To see this, consider the hypothetical scenario of free mobility. In that case, we should not observe discrete changes in famine mortality rates along any administrative border.

Figure 2b plots the mortality residuals from a regression controlling for 1926 Ukrainian population share against distance to the border. There is no border effect once we control for the ethnic Ukrainians rural population share of each district. Conceptually, this accounts for the decline in ethnic Ukrainian share when crossing the border (see Figure 2c). Thus, anti-Ukrainian bias was delineated along ethnic and not administrative lines.

5 Government Policy

5.1 Realized Procurement and Retention

Ukrainian bias in famine mortality is consistent with the presence of anti-Ukrainian bias in policy. In this section, we investigate whether there is bias in the most important policy for food distribution: centrally planned grain procurement. Such positive statistical evidence can help address the lack of conclusive documentary evidence.

Table 4 documents the relationship between Ukrainian population share, centrally planned procurement, retention and famine mortality. The sample size is smaller than the main analysis because of the limited availability of procurement and production data. In column (1), the dependent variable is realized procurement as a share of realized production. We use our corrected grain production. The estimate is similar to equation (1), except that we no longer control for predicted grain and its interaction with famine because grain production is the denominator of the dependent variable. The Ukrainian interaction coefficient is positive, 0.180, and statistically significant at the 1% level. This implies that, all else equal, the share of production taken away from a province that was 100% Ukrainian was 18 percentage-points higher than a province that had no Ukrainians.

Column (2) examines per capita retention (production minus procurement) as the dependent variable. It shows that all else equal, a province with 100% rural Ukrainian population share retained 1.073 kilograms

per capita per day less than a province with no Ukrainians. The Ukrainian interaction coefficient is positive and statistically significant at the 1% level. To examine the timing of anti-Ukrainian bias in food retention, we replace the interaction of rural ethnic Ukrainian population share and the famine dummy with the interaction of rural ethnic Ukrainian population share and year dummy variables. 1928 is the omitted reference group. Figure 1e (and Appendix Table A.6 column 2) shows that Ukrainian population share is uncorrelated with food retention for most years, but negatively correlated in 1932. These estimates show that there was anti-Ukrainian bias in centrally planned food procurement on the eve of the famine.

Columns (3) and (4) document the relationship between mortality and grain retention. In column (3), we regress mortality on realized grain retention and its squared term. Consistent with the positive and concave relationship between food consumption and mortality, we find that the coefficient for grain retention is negative and the coefficient for the squared term is positive. Both estimates are statistically significant at the 5% level. In column (4), we add the interactions of these variables with the famine dummy. The interactions are statistically zero, while the uninteracted terms are similar to column (3). This is a sanity check and reflects the fact that the biological relationship between food consumption and mortality is stable over time. The estimates support the interpretation that the Soviets procured a higher share of 1932 production from Ukrainians, which resulted in lower food retention and higher famine mortality.

To understand the importance of anti-Ukrainian bias in grain procurement policy for famine mortality, we conduct a simple quantification exercise that compares famine mortality when there is bias in retention to the counterfactual of no bias.

First, we estimate the relationship between mortality and grain retention so that we can predict mortality at different levels of retention.

$$mortality_{i,t+1} = F(retention_{it}) + \Gamma X_{it} + \eta_i + \delta_t + \varepsilon_{it}. \tag{2}$$

Equation (2) is conceptually similar to the estimates shown in Table 4, but with a more flexible functional form and fewer controls for parsimony. $F(retention_{it})$ is a flexible step function defined over 0.5 kilogram per capita per day intervals of grain retention, X_{it} includes province characteristics that may affect mortality (e.g., urbanization), and η_i and δ_t are province and year fixed effects. The estimates show that lower retention is associated with higher mortality. The results are robust if in addition to urbanization, we control for urbanization interacted with the famine indicator and the official 1928 grain production interacted with the famine indicator (see Appendix Figure A.3 and Appendix Table A.10).

Second, we estimate famine mortality with anti-Ukrainian bias in grain retention. This is the difference between the number of deaths in 1933 predicted by realized 1932 retention and the number of deaths in non-famine years. The number of deaths predicted by realized retention can be obtained by applying 1932 realized grain retention to the estimates from equation (2) and Table 2 column (2): predicted 1933 mortality is 1.42 million deaths in Ukraine and 4.41 million deaths in the full sample. Average predicted mortality for non-famine years is 0.52 million for Ukraine and 2.72 million for the full sample. Thus, famine mortality with bias is 0.9 million for Ukraine ($1.42 - 0.52 = 0.9$) and 1.7 million for the whole sample ($4.41 - 2.72 = 1.7$).

Next, we estimate famine mortality for the counterfactual of no bias in grain retention. This is the

difference between the number of 1933 deaths predicted by the counterfactual of no bias in retention and the number of deaths in non-famine years. To calculate the counterfactual 1933 mortality, we first predict counterfactual 1932 grain retention. We use the estimates of the relationship between retention and the interactions of Ukrainian population share and year dummy variables shown in Figure 1e. We set Ukrainian share to zero in 1932 to predict counterfactual grain retention for the famine. Then, we apply the counterfactual grain retention to the estimates from equation (2) to predict the counterfactual 1933 mortality: predicted 1933 deaths are 0.9 million in Ukraine and 3.84 million in the full sample. Thus, the counterfactual famine mortality is 0.38 million for Ukraine ($0.9 - 0.52 = 0.38$) and 1.12 million for the whole sample ($3.84 - 2.72 = 1.12$).

Finally, we estimate the contribution of bias in grain retention on famine mortality by comparing famine mortality when there is anti-Ukrainian bias to the counterfactual famine mortality of no bias in retention. It follows that anti-Ukrainian bias in grain retention explains 58% of excess deaths in Ukraine ($1 - 0.38/0.9 = 0.58$) and 34% of all excess deaths in our sample ($1 - 1.12/1.7 = 0.34$). Since overall anti-Ukrainian bias explains 77% and 92% of famine mortality in the whole sample and Ukraine, the estimates imply that approximately half of the total anti-Ukrainian bias effect on famine mortality takes place through bias in grain retention.

The magnitudes are consistent with the importance of grain procurement for determining food availability, but also leave room for other policies to contribute to food availability, such as procurement of potatoes and other non-grain agricultural products, migration restrictions or aid relief.

5.2 Production and Procurement Targets

To understand whether anti-Ukrainian bias in grain procurement was centrally planned and intentional, we examine grain production and procurement targets published before the famine. The First Five-Year Plan, published in 1928, laid out production and procurement targets for each province for the years from 1928 to 1933. In Table 4 Panel B column (5), we regress per capita production targets on the rural share of ethnic Ukrainians. We control for year fixed effects to account for the fact that the planners assumed a high rate of growth in all regions, and officially reported 1928 per capita grain production, which was the measure used by planners to account for regional agricultural productivity. We do not control for province fixed effects because of the limited variation in a five-year panel. The coefficients for Ukrainians and grain 1928 are both positive and statistically significant. For two places with the same observed production in 1928, the one with more ethnic Ukrainians was expected to produce more. This could reflect the regime's desire to be harsher with Ukrainians or a belief that Ukrainian peasants were less productive than other peasants living in similar natural conditions and had more room to expand production.

Column (6) examines per capita procurement targets as the dependent variable, while controlling for production targets, which captures differences in perceived production capacities across regions. For two provinces assigned the same production target, the one with more Ukrainians is assigned a higher procurement target. Column (7) shows a similar pattern when we examine procurement as a share of production as the dependent variable. Thus, the regime intended to take more grain from Ukrainian areas after conditioning for factors such as production capacity. Since we control for production targets, the Ukrainian bias in

grain procurement targets cannot be driven by the belief that Ukrainian grain production had more room to grow.

In column (8), we replace procurement ratios with per capita grain retention targets (the difference between the production and procurement targets) as the dependent variable. The coefficient for ethnic Ukrainian population share is -0.275 and statistically significant at the 1% level. For two places facing the same production targets, central planners intended for the one with 100% ethnic Ukrainians to have 0.275 kilograms of grain per capita per day less than the one with no Ukrainians. The magnitude of the discrepancy is sizable: it is approximately one-third of the official Soviet food requirement (see Section 3).

The target data show that as early as 1928, the regime had planned for Ukrainian areas to retain less grain than other regions with the same level of grain production. However, the data do not imply that the regime planned to kill Ukrainians because the production targets are much higher than actual grain production for all regions and the degree of unequal grain retention implied by the target data would not result in famine mortality at high levels of grain retention. Thus, the results leave open the question of the timing of the decision to let Ukrainians die. In one extreme scenario, the regime aimed to reduce the Ukrainian population size *ex ante* and set procurement targets in 1928 that would lead to high Ukrainian famine mortality. This assumes that the regime secretly knew what true 1932 production would be, but planned to enforce the procurement targets. In an alternative scenario, the regime aimed to penalize Ukrainians for being troublesome by leaving them with less surplus without the intention to cause famine. Stalin was known to have advocated using the over-procurement of food to discipline peasants (Danilov et al., eds, 1999-2006; Davies and Wheatcroft, 2009), and to have rewarded loyal ethnic groups and penalized disloyal ones (Polyan, 2001). In this scenario, the regime intended for Ukrainians to retain less food than others, but planned for Ukrainians to retain enough food for subsistence; when production fell in 1932, the state made the decision to enforce higher procurement from Ukrainians. The truth may also be somewhere in between: the leadership knew that production targets were too optimistic, but did not fully predict the severity of the fall in production in 1932; and faced with the choice of giving up procurement versus letting Ukrainians die, they chose the latter. In either case, the regime is culpable for the high rates of Ukrainian famine mortality. The question is about the timing of the decision to allow Ukrainians to die from famine.

5.3 Motivation of Anti-Ukrainian Bias

As we discussed in the Introduction, the primary political economic explanation of Ukrainian famine mortality is that Ukrainians were repressed because of their importance to Soviet agriculture and their strong resistance to Soviet rural economic policy. It is beyond the scope of this paper to provide conclusive evidence on the motivations of anti-Ukrainian bias. This section provides a speculative discussion and some clues for future research.

Ukrainians had a strong group identity that included their own language and culture, which facilitated collective action. The Ukrainian communist party was the largest national branch of the Soviet Communist Party and viewed itself as representing the interests of ethnic Ukrainians across the Soviet Union, including those who lived outside the boundaries of the republic. Strong political opposition from Ukrainian nationalists during the Civil War was central to the Bolsheviks' "national question" (Graziosi, 2015). The

Bolsheviks were concerned about controlling all peasants, but especially concerned about controlling the Ukrainian peasantry.

We provide two stylized facts consistent with this political-economic motive. The first is to validate the claim that ethnic Ukrainians offered stronger resistance to collectivization than other ethnic groups. We collect data on de-classified secret police reports about peasant resistance to collectivization. We find that the positive slope between collectivization (the share of households that have been collectivized) and resistance is steeper in the ten provinces with Ukrainian population share above the sample median than in the nine provinces with Ukrainian population share below the median, controlling for official 1928 grain production and urban population share. The data show that all else equal, Ukrainian peasants resisted collectivization more intensely.[8] This is consistent with the Bolsheviks' fear that political resistance was stronger from Ukrainians, as well as our finding that anti-Ukrainian bias extended to ethnic Ukrainians across the Soviet Union.

Second, we document that the famine mortality-Ukrainian gradient is steeper in places that were more important for agricultural production. This implies that there was more intense targeting of Ukrainians in agriculturally productive places. We regress mortality on the triple interaction of Ukrainian population share, the famine dummy variable, and the importance of a region for rural economic production as perceived by the state (measured with official 1928 grain production). Table 5 shows that the triple interaction coefficient with 1928 grain production is positive and statistically significant at the 1% level. In column (2), we control for a parsimonious measure of administrative capacity and political zealousness (the first principal component of the share of votes for the Bolshevik Party in the 1917 Constituency Assembly election, the size of the Communist Party averaged over 1922, 1927 and 1931, and the number of the 1930 Party Congress delegates; we also examine these variables separately in Appendix Section D). The triple interaction coefficient with 1928 grain production is robust to the additional control, while the triple interaction coefficient with the administrative capacity proxy is small in magnitude and statistically insignificant. Thus, the Ukrainian-mortality gradient is steeper in places that are more important for agriculture, but does not vary with these other factors that are unrelated to agriculture. Column (3) repeats the estimate with district-level data. The triple interaction of Ukrainian population share, suitability for grain cultivation and the famine dummy is positive, large in magnitude and statistically significant at the 1% level. As we discussed earlier in the paper, finding similar patterns at different administrative levels supports the interpretation that the patterns we observe are due to centrally planned policy.

Additional Mechanisms There are several complementary hypotheses for why the Soviet regime repressed Ukrainians for political economic reasons. A variation of the hypothesis discussed above stems from the well-known insights of Horowitz (1985). The logic is as follows. Given the need to control agriculture, the regime targeted ethnic Ukrainians because it lacked more precise information on the likelihood of subversion. Thus, ethnic Ukrainian population share was used as a crude marker on which central planners conditioned policies.

[8] We plot the fitted lines and residuals in Appendix Figures A.4a and A.4b. The p-value for the statistical difference between the two slopes is 0.09.

Another hypothesis is that anti-Ukrainian bias is partly driven by asymmetric information and rigidities in the command economy. Meng et al. (2015) first made this argument for the Chinese famine, which occurred in a centrally planned procurement system modeled after the Soviet one. They argue that the famine was partly due to the government's inability to adjust procurement because it did not trust local information. Farmers are not the residual claimants of production and therefore incentivized to under-report production or shirk. Local bureaucrats may over-report production to curry favor with the Party or under-report to build a local power base. Both are bad for the regime. Over-reporting can lead to over-procurement and famine, which can lower future production and be politically destabilizing. Under-reporting lowers revenues. Thus, the government will not lower procurement until they can verify that the production drop is not caused by peasants shirking (or if famine would be too politically costly). Information rigidities can explain the anti-Ukrainian bias in Soviet policy and famine mortality if the Bolsheviks were particularly distrustful of Ukrainians. This theory of famine is consistent with the Bolsheviks' primary objective to control agricultural production and under-reporting production can be viewed as a form of Ukrainian resistance that threatens the Bolsheviks.

Note that there are important differences in the political economic contexts of the two famines. The Chinese famine occurred in the rural political power base of the Chinese Communist Party, including the home provinces of the top Party leaders. In contrast, by the time of the Ukrainian famine, there had been a history of political conflict between the Ukrainian peasantry and the Bolsheviks, whose political power base lie in urban areas. Agricultural production was also more important to the Chinese economy in 1959 (approximately 80% of GDP) than the Soviet economy in 1928 (approximately 50% of GDP).

The variants of the political economic motivation discussed in this section are not mutually exclusive; nor do they preclude other motivations for anti-Ukrainian bias.

6 Rural population and Grain Production after the Famine

This section examines long-run economic and demographic patterns in areas that had a higher share of Ukrainian population and suffered disproportionately high famine mortality rates.

First, we use all available population censuses from 1897 to 2002 to examine total rural population. We regress log rural population size on the interaction of ethnic Ukrainian share and (census) year fixed effects, controlling for province and year fixed effects. Figure 3a (Appendix Table A.11 column 1) presents the interaction coefficients and its 95% confidence intervals. The figure shows that relative to areas with no Ukrainians, the rural populations in Ukrainian areas declined in size immediately after the famine and had not recovered by 1939. Rural population had recovered by 1959 in the first population census after the war.

In Figure 3b, we examine the rural population share of ethnic Ukrainians and ethnic Russians. The estimates are imprecise and should be interpreted as suggestive. Nevertheless, the patterns are striking: the share of ethnic Ukrainians declined after the famine and the decline was permanent. This goes against the concern that famine mortality in regions with high ethnic Ukrainian population share was driven by the deaths of other ethnicities and not that of ethnic Ukrainians ("ecological fallacy"). The statistical evidence is consistent with the fact that there are no known accounts of such mortality patterns.

Second, we examine grain production. As with the accounting exercise, we use corrected grain figures for the years 1928-1933 and official grain figures for all other years. We regress log total grain production for each province and year on the interaction of pre-famine ethnic Ukrainian population share and year fixed effects, controlling for the time-invariant measure of suitability for grain cultivation interacted with year fixed effects, 48 monthly temperature and precipitation variables for years t and $t-1$, and province and year fixed effects. Figure 3c (Appendix Table A.11 Column 4) presents the interaction coefficients of interest. Consistent with the republic-level accounting exercise, we observe a larger production drop in Ukrainian areas. In addition, we observe that production in these areas remained lower than others for a total of four years and had recovered by 1936. Thus, production in famine-stricken areas recovered before the labor supply.

Agriculture in the early 1930s still mostly relied on traditional means of production and a key feature of Soviet economic policy was to mechanize agriculture. The most important mode of mechanization was the adoption of tractors, which were centrally allocated by Moscow. Thus, we examine the allocation of tractors to understand whether the regime increased mechanization in famine-stricken regions to boost production. We collect archival data on province-level tractors allocation for 1927 to 1939. In normal times, tractors were allocated based on sown area and the importance of a location for agricultural production. Thus, we regress total tractor horse power per hectare of 1928 sown area on the interaction of year dummy variables and the pre-famine Ukrainian population share, controlling for the time-invariant measure of suitability for grain production interacted with year fixed effects, and province and year fixed effects. Figure 3d (Appendix Table A.11 Column 5) shows small increases in the number of tractors allocated to regions that had a high pre-famine Ukrainian population share during 1927 to 1932, and a large increase during 1933 to 1939. Since the allocation of tractors may have been accompanied by other inputs such as better fertilizers, we interpret tractors as broadly reflecting the use of advanced inputs or the mechanization of agriculture. These estimates together with the results on rural population and production suggest that tractors were used to moderate the loss of labor and boost production in famine-stricken regions.

7 Conclusion

The deaths of approximately 2.1 to 3.15 million ethnic Ukrainians during the 1932-33 Soviet Famine is one of the most controversial human tragedies of the 20th century. Ethnic Ukrainians in the Soviet Union declined 21.3% to 16.5% of the total population between 1926 and 1939. In areas that the Bolshevik regime marked as important for grain production, ethnic Russians replaced ethnic Ukrainians as the largest ethnic group. This paper provides the first systematic evidence that disproportionately high Ukrainian famine deaths were not an unintended consequence, but an outcome of anti-Ukrainian bias in Soviet policy. The Bolsheviks systematically over-procured food from ethnic Ukrainians, province by province, district by district, within and outside of Ukraine. Anti-Ukrainian bias in the grain distribution policy and migration restrictions both contributed to high Ukrainian famine mortality.

We hope that the rigorous empirical evidence provided in this paper will establish a ground truth for future research and public discussions.

Our findings suggest several interesting questions for future research. One is to better understand how other policies contributed to Ukrainian mortality. Another is to understand the political-economic tradeoffs of mass repression for the regime. The existing evidence is ambivalent in our context. On the one hand, rural production recovered within a few years of the famine and the Bolsheviks remained in power for another fifty years. On the other hand, the famine led to Ukrainian animosity towards Russians that lasted generations (Korovkin and Makarin, 2023; Rozenas and Zhukov, 2019). Relatedly, it is interesting to understand the role that the repression had in perpetuating ethnic conflict. Prior to the famine, there was little hostility between ethnic Ukrainians and ethnic Russians. Did the economically motivated repression inadvertently deepen ethnic division? The fact that Russian officials removed the commemorative memorial of the victims of the *Holodomor* to combat "political disinformation" when they occupied Mariupol in 2022 highlights the importance of the famine for Ukrainians today.[9]

References

Andreev, E.M., L.E. Darskij, and T.L. Kharkova, *Naselenie Sovetskogo Soyuza: 1922-1991*, Moscow: Nauka, 1993.

_ , _ , and _ , *Demographicheskaya Istroriya Rossii: 1927-1959*, Moscow: Informatika, 1998.

Applebaum, Anne, *Red Famine: Stalin's War on Ukraine*, New York: Penguin Press, 2017.

Becker, Sascha O., Sharun Mukand, and Ivan Yotzov, "Persecution, pogroms and genocide: A conceptual framework and new evidence," *Explorations in Economic History*, 2022, *86*, 101471.

Berelowitch, A. and V. Danilov, eds, *Les campagnes Soviétiques vues par la Tchéka–O.G.P.U.–N.K.V.D. 1918–1939. Recueil de documents et matériaux en 4 volumes. Sovetskaya derevnya glazami VChK–OGPU–NKVD. 1918–1939 gg. Dokumenty i materialy v 4 tomakh*, Moscow: Rosspen, 2000-2012.

Bogoyavlensky, Dmitry, "O pripiskakh v perepisi 1939 g [On the centralized additions in the 1939 census]," *Demoscope.ru*, 2013.

Buggle, Johannes C and Ruben Durante, "Climate Risk, Cooperation and the Co-Evolution of Culture and Institutions," *The Economic Journal*, 01 2021, *131* (637), 1947–1987.

Bukhman, K., "Golod 1921 goda i deyatelnost inostrannykh organizatsiy [Famine of 1921 and the activities of foreign organizations]," *Vestnik statistiki*, 1923, (14), 87–113.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller, "Bootstrap-Based Improvements for Inference with Clustered Errors," *The Review of Economics and Statistics*, 08 2008, *90* (3), 414–427.

Canay, Ivan A., Andres Santos, and Azeem M. Shaikh, "The Wild Bootstrap with a "Small" Number of "Large" Clusters," *The Review of Economics and Statistics*, 05 2021, *103* (2), 346–363.

Castañeda Dower, Paul and Andrei Markevich, "Labor Misallocation and Mass Mobilization: Russian Agriculture during the Great War," *The Review of Economics and Statistics*, 2018, *100(2)*, 245–59.

[9] See RIA Novosti channel in telegram, October 19, 2022, https://t.me/rian_ru/182454. Accessed October 24, 2022.

_ **and** _ , "Did Industrialization Increase Support for the Radical Left? Evidence from the 1917 Russian Revolution," *Working paper*, 2023.

Charnysh, Volha, "Explaining Out-Group Bias in Weak States: Religion and Legibility in the 1891/1892 Russian Famine," *World Politics*, 2022, *74* (2), 205–248.

Cheremukhin, Anton, Mikhail Golosov, Sergei Guriev, and Aleh Tsyvinski, "The Industrialization and Economic Development of Russia through the Lens of a Neoclassical Growth Model," *The Review of Economic Studies*, 2017, *84* (2), 613–649.

Colella, Fabrizio, Rafael Lalive, Seyhun Orcan Sakalli, and Mathias Thoenig, "Inference with Arbitrary Clustering," 2019.

Conquest, Robert, *The Harvest of Sorrow: Soviet Collectivization and the Terror-Famine*, Oxford, UK: Oxford University Press, 1986.

Danilov, Victor, Robert Manning, and Lynne Viola, eds, *Tragediya Sovetskoj Derevni. Kollektivizatsiya i raskulachivanie. Dokumenti i materialy v 5 tomakh, 1927-1939*, Rosspen, 1999-2006.

Davies, R. W., Oleg V. Khlevniuk, E. A. Rees, Liudmila P. Kosheleva, and Larisa A. Rogovaya, eds, *The Stalin-Kaganovich Correspondence, 1931-36*, New Haven: Yale University Press, 2003.

Davies, Robert W. and Stephen G. Wheatcroft, *The Years of Hunger: Soviet Agriculture, 1931-1933. Second edition.*, Vol. 5 of *The Industrialization of Soviet Russia*, Basingstoke, UK: Palgrave Macmillan, 2009.

Departament tamozhennykh sborov, ed., *Obzor vneshney torgovli Rossii po Yevropeyskoy i Aziatskoy granitsam za 1913 g* 1914.

Dyson, Tim and Cormac Ó Gráda, eds, *Famine Demography: Perspectives from the Past and Present*, Oxford, UK: Oxford University Press, 2002.

Egorov, Georgy and Konstantin Sonin, "The Political Economics of Non-democracy," *University of Chicago, Becker Friedman Institute for Economics Working Paper*, 2020, (2020-142).

Ellman, Michael, "The Role of Leadership Perceptions and of Intent in the Soviet Famine of 1931-1934," *Europe-Asia Studies*, 09 2005, *57* (6), 823–841.

Esteban, Joan, Massimo Morelli, and Dominic Rohner, "Strategic Mass Killings," *Journal of Political Economy*, 2015, *123(5)*, 1087–1132.

Gokmen, Gunes and Dmitrii Kofanov, "Career incentives in political hierarchy: evidence from Imperial Russia," *European Review of Economic History*, 2020, *24(2)*, 264–287.

Graziosi, Andrea, "The uses of hunger Stalin's solution of the peasant and national questions in Soviet Ukraine, 1932 to 1933," in Lubomyr Luciuk Declan Curran and Andrew G. Newby, eds., *Famines in European History. The Last Great European Famines Reconsidered*, Routledge, 2015, pp. 223–260.

Heldring, Leander, "The Origins of Violence in Rwanda," *The Review of Economic Studies*, 06 2020.

Horowitz, Donald L., *Ethnic groups in conflict*, Berkeley: University of California Press, 1985.

Jha, Saumitra, "'Unfinished business': Historic complementarities, political competition and ethnic violence in Gujarat," *Journal of Economic Behavior & Organization*, 2014, *104*, 18–34.

Khrushchev, N.S., *Nikita S. Khrushchev: Vremia, Liudi, Vlast* 1970.

Kondrashin, Viktor, *Golod 1932-1933 godov: Tragediya Sovetskoj derevni*, Moscow: Rosspen, 2008.

Korovkin, Vasily and Alexey Makarin, "Conflict and Intergroup Trade: Evidence from the 2014 Russia-Ukraine Crisis," *American Economic Review*, January 2023, *113* (1), 34–70.

Kotkin, Stephen, *Stalin. Waiting for Hitler, 1929-1941*, New York: Penguin Press, 2017.

Levchuk, Nataliia, Oleh Wolowyna, Omelian Rudnytskyi, Alla Kovbasiuk, and Natalia Kulyk, "Regional 1932-1933 Famine Losses: A Comparative Analysis of Ukraine and Russia," *Nationalities Papers*, 2020, *48* (3), 492–512.

Lositskiy, A. E., "Pishchevoe znachenie sovremennikh produktov pitaniya [The nutritional value of modern foods]," *Vestnik statistiki*, 1920, (1-4), 102–121.

_ **, ed.**, *Sostoyane pitaniya gorodskogo naseleniya SSSR 1919–1924 gg*, Vol. XXX, Issue 1 of *Trudy TsSU* 1926.

_ **, ed.**, *Sostoyane pitaniya sel'skogo naseleniya SSSR 1920–1924 gg*, Vol. XXX, Issue 2 of *Trudy TsSU* 1928.

Markevich, Andrei, Natalya Naumenko, and Nancy Qian, "The Causes of Ukrainian Famine Mortality, 1932-33," Working Paper 29089, National Bureau of Economic Research May 2023.

Martin, Terry, *The Affirmative Action Empire: Nations and Nationalism in the Soviet Union, 1923-1939*, Cornell University Press, 2001.

Matsuura, Kenji and Cort Willmott, "Terrestrial Air Temperature and Precipitation: 1900-2014 Gridded Monthly Time Series (Version 4.01)," *National Oceanic and Atmospheric Administration*, 2014.

Meng, Xin, Nancy Qian, and Pierre Yared, "The Institutional Causes of China's Great Famine, 1959-1961," *The Review of Economic Studies*, 2015, *82* (4), 1568–1611.

Meslé, France, Jacques Vallin, and Evgeny Andreev, "Demographic Consequences of the Great Famine: Then and Now," in Andrea Graziosi, Lubomyr A. Hajda, and Halyna Hryn, eds., *After the Holodomor: The Enduring Impact of the Great Famine on Ukraine*, Harvard University Press, 2013.

Murin, Yuri, ed., *Pisatel' i vozhd': perepiska M.A. Sholokhova s I.V. Stalinym 1931-1950 gody : sbornik dokumentov iz lichnogo arkhiva I.V. Stalina*, Raritet, 1997.

Naumenko, Natalya, "The Political Economy of Famine: the Ukrainian Famine of 1933," *The Journal of Economic History*, 2021, *81* (1), 156–197.

_ , "Collectivization, the Industrialization Debate, and the First-Five Year Plans," in A.V. Gevorkyan, ed., *The Oxford Handbook of Post-Socialist Economies in Europe, Caucasus, and Central Asia*, Oxford University Press, forthcoming.

Ó Gráda, Cormac, *Famine: A Short History*, Princeton, NJ: Princeton University Press, 2009.

Polyakov, Yu.A. and V.B. Zhiromskaya, *Naselenie Rossii v XX veke: istoricheskie ocherky. Vol. 1 1900-1939*, Rosspen, 2000.

Polyan, Pavel, *Ne po svoej vole... Istoriya i geografiya prinuditelnikh migratsij v SSSR*, Moscow: O.G.I, 2001.

Protasov, L.G., A.V. Repnikov V.V. Zhuravlev, and V.V. Shelokhaev, *Vserossijskoe uchreditelnoe sobranie. Entsiklopedia. (All-Russian Constitutional Assembly. Encyclopedia)*, Moscow: Rosspen, 2014.

Rogall, Thorsten, "Mobilizing the Masses for Genocide," *American Economic Review*, January 2021, *111* (1), 41–72.

Rozenas, Arturas and Yuri M. Zhukov, "Mass Repression and Political Loyalty: Evidence from Stalin's 'Terror by Hunger'," *American Political Science Review*, 2019, *113* (2), 569–583.

Rudnytskyi, Omelian, Natalia Levchuk, Oleh Wolowyna, Pavlo Shevchuk, and Alla Kovbasiuk, "Demography of a man-made human catastrophe: The case of massive famine in Ukraine 1932-1933," *Canadian Studies in Population*, 2015, *42* (1-2), 53–80.

Sen, Amartya, *Poverty and Famines. An Essay on Entitlement and Deprivation*, Oxford, UK: Oxford University Press, 1981.

State Duma, "Russian: Postanovleniye GD FS RF ot 02.04.2008 N 262-5 GD O Zayavlenii Gosudarstvennoy Dumy Federal'nogo Sobraniya Rossiyskoy Federatsii "Pamyati zhertv goloda 30-kh godov na territorii SSSR"," 2008.

Tauger, Mark B., "The 1932 Harvest and the Famine of 1933," *Slavic Review*, 1991, *50* (1), 70–89.

TsSU, *Vsesoyuznaya perepis naseleniya 1926 goda. Vol. 1–41*, TsSU, 1928-1930.

TsSUNKhU, *Skhematicheskie karty rajonov SSSR*, Moscow: TsSUNKhU Gosplana SSSR, 1934.

Viola, Lynne, *Peasant Rebels Under Stalin: Collectivization and the Culture of Peasant Resistance*, Oxford, UK: Oxford University Press, 1996.

_ , *The Unknown Gulag: The Lost World of Stalin's Special Settlements*, Oxford: Oxford University Press, 2007.

Wheatcroft, Stephen G. and Anthony Garnaut, "Poteri naseleniia v otdelnykh raionakh SSSR (1929-1934): Statistika, karty i sravnitelnyi analiz (osoboie polozhenie Ukrainy)," in "Holod 1993: Ukraintsi," Kyiv: Ukrainian Institute of National Memory, 2013, pp. 376–391.

Yanagizawa-Drott, David, "Propaganda and Conflict: Evidence from the Rwandan Genocide," *The Quarterly Journal of Economics*, 11 2014, *129* (4), 1947–1994.

Zelenin, I. E., N. A. Ivnitskii, V. V. Kondrashin, and E. N.. Oskolkov, *O golode 1932-1933 godov i ego otsekne na Ukraine*, Otechestvennaya istoriya 6, 1994.

Zhuravskaya, Ekaterina, Sergei Guriev, and Andrei Markevich, "New Russian Economic History," *The Journal of Economic Literature*, forthcoming.

Table 1: Grain Availability

	1927	1928	1929	1930	1931	1932	1933	1937	1939
	I. Ukraine								
(1) Total population (mln)	29.0	29.6	30.3	30.8	31.3	31.7	31.9	28.4	29.6
(2) Rural population (mln)	23.6	24.6	24.9	25.1	25.0	24.8	25.0	18.8	18.7
(3) Production (mln tons)	.	14.9	18.7	23.2	16.8	9.1	16.1	.	.
(4) Procurement (mln tons)	4.0	2.0	5.3	7.7	7.3	4.2	6.1	.	.
(5) Rural retention (kg/person/day)	.	1.436	1.477	1.693	1.042	0.538	1.097	.	.
(6) Rural retention, no procurement (kg/person/day)	.	1.662	2.060	2.531	1.838	1.005	1.768	.	.
(7) Rural and urban retention, no procurement (kg/person/day)	.	1.379	1.696	2.065	1.468	0.786	1.383	.	.
(8) Food needs for heavy labor (kg/person/day)	0.78	0.79	0.78	0.78	0.78	0.78	0.78	0.76	0.75
	II. USSR – no Ukraine								
(9) Total population (mln)	118.0	120.8	124.0	126.7	129.3	131.5	133.8	133.7	135.9
(10) Production (mln tons)	.	58.5	53.1	60.1	47.8	48.0	51.2	.	.
(11) Rural and urban retention, no procurement (kg/person/day)	.	1.326	1.172	1.299	1.012	1.001	1.048	.	.
(12) Food needs for heavy labor (kg/person/day)	0.78	0.78	0.78	0.78	0.78	0.77	0.77	0.76	0.76

Notes: Data for population and procurement are official statistics. Production is revised by the authors using archival sources. See the text and Appendix for details. Retention is the difference between production and procurement. Food needs are calculated by the authors and take into account the demographic composition (e.g., age, gender, rural/urban) as reported in the population censuses. They are based on official guidelines for maximum caloric needs for each group as reported by Lositskiy, ed (1926, 1928). Panel I includes Ukraine. Panel II includes all other republics in the USSR.

Table 2: Famine Mortality and Ethnic Ukrainian Population Share – Province-level Estimates

	Dependent Variable:									
	Mortality in Year $t+1$								Natality in Year $t+1$	Cohort Size in Year $t+1$
	Control for Corrected Grain Production	Baseline	Kulaks × Famine	Livestock Change× Famine	Omit Province FE, control for Ukrainians	Omit Ukraine	Other Minorities× Famine	1892 Famine		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ukrainians× Famine	0.050	0.051	0.054	0.053	0.051	0.086	0.049	-0.0002	-0.014	-0.009
	(0.002)	(0.006)	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)	(0.003)	(0.003)	(0.001)
Standardized Coef.	*0.813*	*0.826*	*0.870*	*0.858*	*0.831*	*0.840*	*0.790*	*-0.004*	*-0.432*	*-0.608*
Grain	0.001	0.0002	0.0002	0.0002	-0.0002	0.0003	0.0002	0.116	0.001	0.00003
	(0.001)	(0.0001)	(0.0002)	(0.0002)	(0.0004)	(0.0001)	(0.0001)	(0.110)	(0.0002)	(0.0001)
Grain× Famine	0.007	0.0004	-0.003	-0.001	-0.0001	-0.003	0.001	0.435	-0.001	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.948)	(0.001)	(0.0003)
Standardized Coef.	*0.275*	*0.023*	*-0.137*	*-0.067*	*-0.005*	*-0.183*	*0.031*	*0.046*	*-0.047*	*-0.120*
Additional Control (column heading)			1.805	0.016	-0.007		-0.007			
			(1.367)	(0.015)	(0.002)		(0.004)			
Observations	107	337	337	337	337	319	337	1,297	337	1,296
R-squared	0.838	0.776	0.791	0.784	0.420	0.758	0.780	0.864	0.820	0.768
Ukrainians + Ukrainians × Famine: Coef.					0.044					
p-val.					<0.001					
Dep. var. mean, non-famine years	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.031	0.039	0.025
Dep. var. mean, famine year	0.031	0.031	0.031	0.031	0.031	0.029	0.031	0.040	0.028	0.017

Notes: The sample includes Ukraine, Russia and Belarus. Observations are at the province and year level. In column (1), the sample includes years from 1928 to 1933; in columns (2)-(7) and (9), the sample includes years from 1922 to 1939; in column (8), the sample includes years from 1885 to 1913; in column (10), the sample includes years from 1922 to 1937. Mortality is the number of deaths divided by the total population. Natality is the number of live births divided by the total population. Cohort size is the birth cohort size reported by the 1939 census divided by the total 1939 population. Ukrainians is the 1926 rural Ukrainian population share. In columns (1)-(7) and (9)-(10), famine is an indicator that equals one in 1932 and zero otherwise; in column (8), famine is an indicator that equals one in 1891 and zero otherwise. All estimates control for per capita grain and grain × Famine; in column (1), grain is production estimates revised by the authors using archival data (see text), in columns (2)-(10) grain is per capita grain production predicted by exogenous factors. Columns (1)-(9) control for urbanization (the urban population share) and urbanization × Famine, column (10) controls for urbanization 1926 × Famine. In addition, column (3) controls for the number of kulak households exiled during 1930-31 per 1930 population interacted with the famine indicator. Column (4) controls for the drop in livestock (horses and cattle) per capita between 1929 and 1931 interacted with the famine indicator. In column (7), "other minorities" include all ethnic groups except Russians and Ukrainians, so the omitted category is Russians. Columns (1)-(4) and (6)-(10) control for province and year FE; column (5) controls for year FE. The standard errors in parentheses are adjusted for spatial correlation within 1,500 km.

Table 3: Famine Mortality and Ethnic Ukrainian Population Share — District-level Estimates

	Dependent Variable: Mortality				
			Monthly Temp and Rain Deviations for Year t-1, t	Omit Ukraine	Other Minorities
	(1)	(2)	(3)	(4)	(5)
Ukrainians × Famine	0.034	0.024	0.024	0.024	0.026
	(0.006)	(0.007)	(0.007)	(0.008)	(0.007)
	[0.006]	[0.007]	[0.007]	[0.008]	[0.007]
Standardized Coef.	*0.438*	*0.311*	*0.311*	*0.209*	*0.334*
Other Minorities × Famine					0.010
					(0.007)
					[0.007]
Observations	3,274	3,274	3,274	2,498	3,274
R-squared	0.797	0.812	0.812	0.783	0.813
District FE	Y	Y	Y	Y	Y
Year FE	Y	N	N	N	N
Province-Year FE	N	Y	Y	Y	Y
Ukrainians					
Mean	*0.255*	*0.255*	*0.255*	*0.070*	*0.255*
Std. Dev.	*0.374*	*0.374*	*0.374*	*0.171*	*0.374*
Dep. var. mean in 1928	*0.019*	*0.019*	*0.019*	*0.020*	*0.019*
Dep. var. mean in 1933	*0.039*	*0.039*	*0.039*	*0.031*	*0.039*

Notes: The sample includes Ukraine and Russia. Observations are at the district and year level; the sample includes two years – 1928 and 1933. Mortality is the number of deaths divided by the total population. Ukrainians is the 1926 rural Ukrainian population share. Famine is an indicator that equals to one in 1933 and zero otherwise. All regressions control for grain suitability × famine, urbanization and urbanization × famine. Columns (1)-(2), (4)-(5) control for monthly temperature and precipitation in years t and $t-1$ (48 additional controls). Column (3) controls for deviations in monthly temperature and precipitation from the long-term (1900-1950) means in years t and $t-1$ (48 additional controls). In column (5), "other minorities" include all ethnic groups except Russians and Ukrainians. The standard errors in parentheses are adjusted for spatial correlation within 400 km. Standard errors clustered at the district level are presented in square brackets.

Table 4: Procurement, Retention and Mortality

	Dependent Variable: Production, Procurement, Retention (kg/person/day) and Mortality							
	A. Realized Amounts				B. Centrally Planned Targets			
	Procurement Share = Procurement/ Production	Retention = Production - Procurement	Mortality in Year $t+1$		Production	Procurement	Procurement Share = Procurement/ Production	Retention = Production - Procurement
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ukrainians × Famine	0.180	-1.073						
	(0.017)	(0.186)						
Ukrainians					0.776	0.275	0.165	-0.275
					(0.191)	(0.071)	(0.020)	(0.071)
Retention			-0.027	-0.026				
			(0.012)	(0.010)				
Retention^2			0.008	0.008				
			(0.004)	(0.003)				
Retention × Famine				-0.016				
				(0.027)				
Retention^2 × Famine				0.004				
				(0.009)				
Official Grain Production 1928					0.907			
					(0.052)			
Production Target						0.540	0.285	0.460
						(0.035)	(0.016)	(0.035)
Observations	107	107	107	107	90	90	90	90
R-squared	0.937	0.781	0.627	0.637	0.691	0.814	0.844	0.723
Dep. var. mean except 1932	*0.222*	*1.018*	*0.021*	*0.021*	*1.648*	*0.257*	*0.085*	*1.390*
Dep. var. mean in 1932	*0.254*	*0.918*	*0.031*	*0.031*	*1.798*	*0.323*	*0.106*	*1.474*

Notes: The sample includes Ukraine, Russia and Belarus. Observations are at the province and year level for the years from 1928 to 1933. Ukrainians is the 1926 rural Ukrainian population share. Panel A: Production is revised by the authors based on archival documents (see text). Retention is measured in kilograms per person per day. Mortality is the number of deaths divided by the total population. All estimates control for urbanization, urbanization × famine, and province and year fixed effects. The standard errors in parentheses are adjusted for spatial correlation within 1,500 km. Panel B: Official grain production 1928, production, procurement and retention targets are measured in kilograms per person per day. All estimates control for year fixed effects. Huber-White robust standard errors are in parentheses.

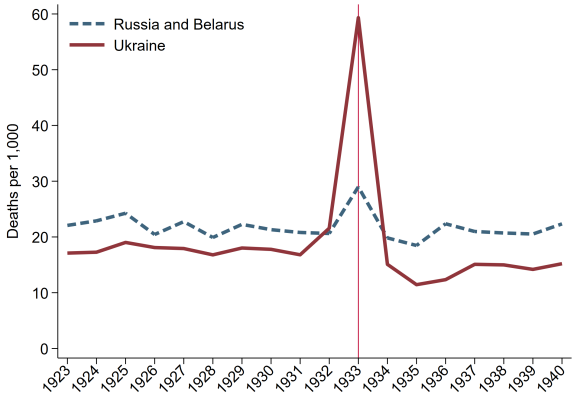
Table 5: Heterogeneous Effects of Grain Productivity on Famine Mortality in Ethnic Ukrainian Areas

	Dependent Variable: Mortality in Year $t+1$			
	Province-Level			District-Level
	(1)	(2)		(3)
Ukrainians × Official Grain Production 1928 × Famine	0.300	0.263	Ukrainians × Grain Suitability × Famine	0.070
	(0.051)	(0.081)		(0.022)
Ukrainians × Famine	-0.174	-0.065	Ukrainians × Famine	-0.010
	(0.035)	(0.156)		(0.017)
Official Grain Production 1928 × Famine	0.0002	0.001	Grain Suitability × Famine	0.005
	(0.002)	(0.002)		(0.006)
Ukrainians × Admin. Capacity × Famine		0.022		
		(0.027)		
Admin. Capacity × Famine		-0.001		
		(0.001)		
Observations	337	337	Observations	3,274
R-squared	0.845	0.846	R-squared	0.784

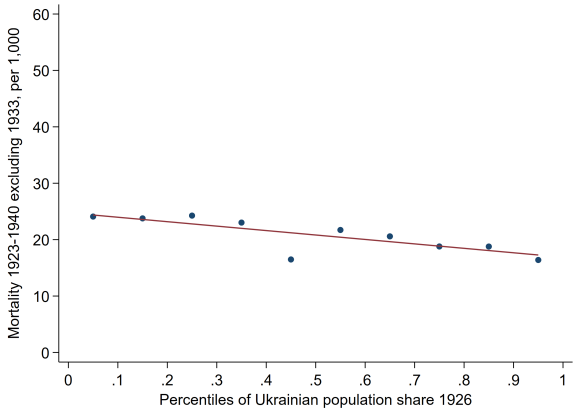
Notes: The province sample includes Ukraine, Russia and Belarus; the distinct sample includes Ukraine and Russia. Cols (1) and (2) control for urbanization, urbanization × famine, Ukrainians × urbanization × famine, grain (per capita grain production predicted by exogenous factors), grain × famine, Ukrainians × grain × famine; and province and year fixed effects. In col (2), Admin. Capacity is the first principal component of the share of votes for the Bolshevik Party in the 1917 Constituency Assembly election, the number of Communist Party Members per 1,000 individuals in each province, and the number of delegates at the 1930 Party Congress. In cols (1) and (2), the standard errors in parentheses are adjusted for spatial correlation within 1,500 km. Col (3) controls for urbanization, urbanization × famine, Ukrainians × urbanization × famine, and district and province-year fixed effects. The standard errors in parentheses are adjusted for spatial correlation within 400 km.

Figure 1: Famine over Time

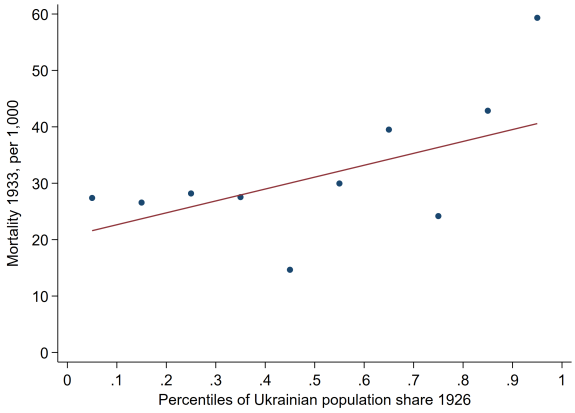
(a) Mortality



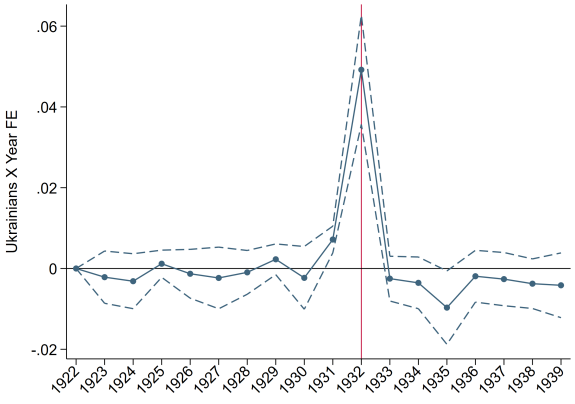
(b) Mortality vs Ukrainian Pop Share



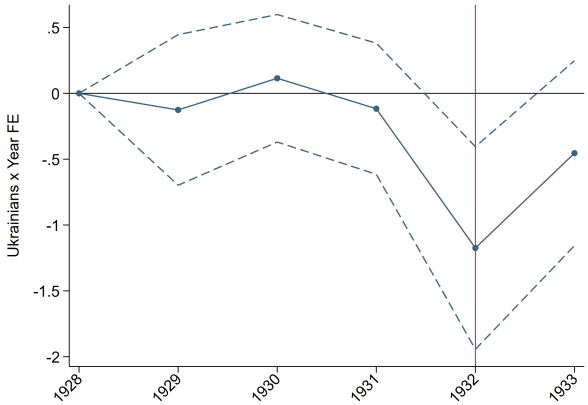
(c) Mortality vs Ukrainian Pop Share – Famine



(d) Coefs. of *Ukrainians* × *Year FEs* on Mortality



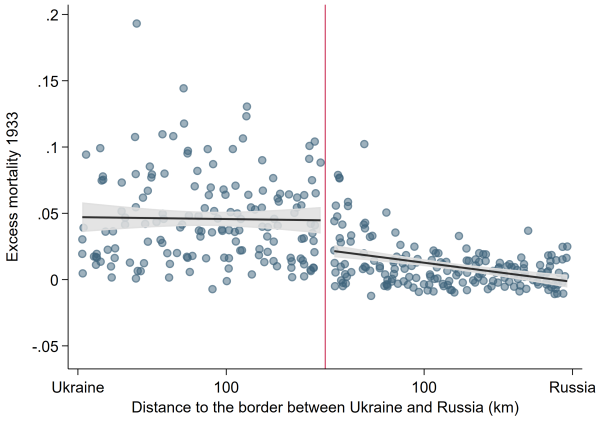
(e) Coefs. of *Ukrainians* × *Year FEs* on Retention



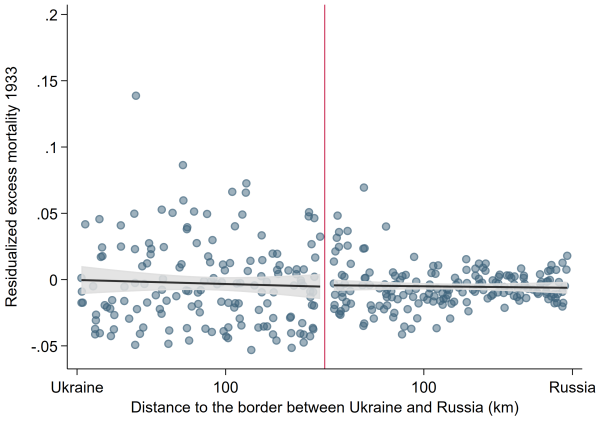
Notes: Mortality is the number of deaths divided by the total population. Figures 1b, 1c show binned scatterplots of provinces' Ukrainian population share (in percentiles) and mortality; Figure 1b uses all years from 1923 to 1940 except 1933, and Figure 1c uses only 1933 mortality. Figures 1d, 1e plot the interaction coefficients of Ukrainian population share and year dummy variables with their 95% confidence intervals. The coefficients and their standard errors are presented in Appendix Table A.6.

Figure 2: Famine Mortality at the Ukrainian-Russian Border

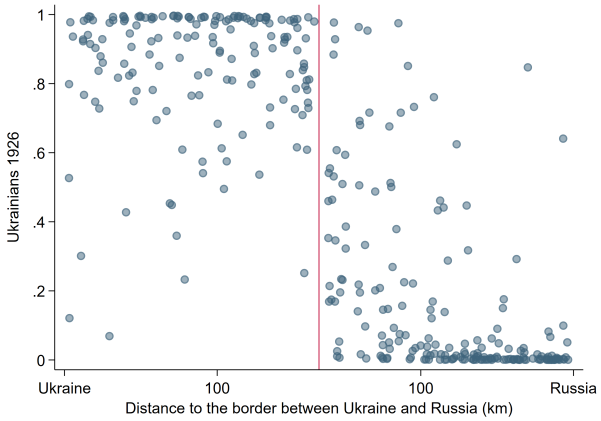
(a) Mortality



(b) Demeaned by 1926 Ukrainian Population Share



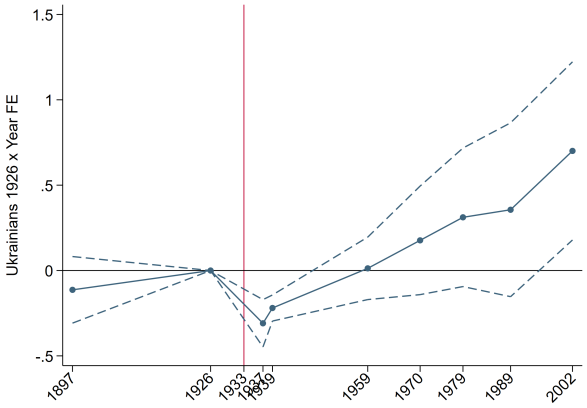
(c) Ukrainian Population Share



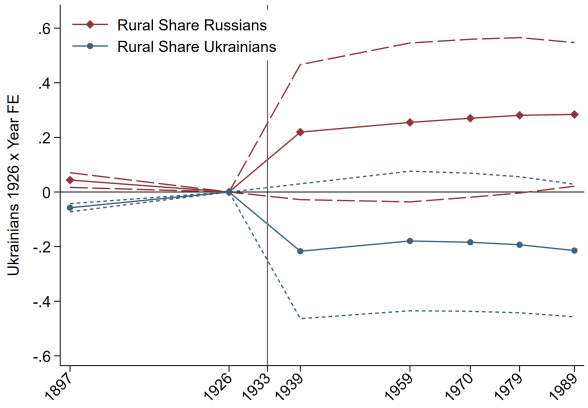
Notes: The figures plot excess mortality against distance to the Ukraine-Russia border and the fitted lines and their 95% confidence intervals for each district in Ukraine and Russia. Excess mortality is the difference between 1933 and 1928 mortality for each district. In Figure 2b, excess mortality is demeaned by the share of ethnic Ukrainians in each district. The estimated magnitude of the border effect is presented in Appendix Table A.9.

Figure 3: Rural Population, Grain Production, Mechanization After the Famine

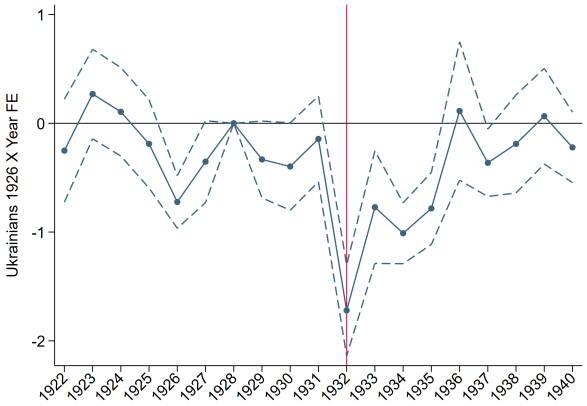
(a) Log Rural Population



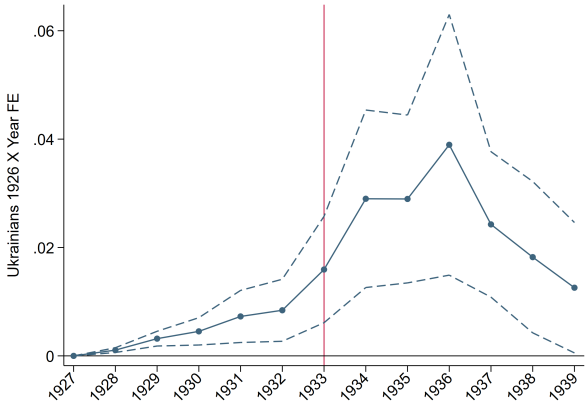
(b) Rural Ethnic Ukrainian and Russian Population Shares



(c) Log Grain Production



(d) Tractors (Tractor Horse Power per Hectare of 1928 Sown Area)



Notes: All figures plot the interaction coefficients of 1926 Ukrainian population share and year dummy variables in a regression that controls for province and year fixed effects. Figure 3c also controls for grain suitability interacted with year fixed effects and monthly temperature and precipitation for years t and $t-1$. Figure 3d also controls for grain suitability interacted with year fixed effects. Appendix Table A.11 reports the coefficients and their standard errors. Figures 3a and 3b use population censuses. See the text for data sources of other figures.

Online Appendix (Not for Publication)

A Revising Grain Production

Appendix Table A.2 presents the aggregate grain counting exercise for Russia and the entire Soviet Union that are analogous to what is shown for Ukraine in Table 1. Panel I presents data for Russia. Panel II for all of the Soviet Union. We also compare our production corrections to those from Davies and Wheatcroft (2009), which are only available for the aggregate level and only as a range (henceforth, DW range).[1]

For brevity, we focus our discussion on the different corrections in Panel II. We approximate "true" production by using grain procurement ratios (procurement as a share of production) reported by the State Planning Committee (Gosplan), and procurement levels. Procurement stocks were directly observed and counted by the government, and the accuracy of procurement ratios reported in the Gosplan has not been disputed. The Gosplan report was a candid evaluation of the First Five-Year Plan shown only to the highest ranking Soviet officials. The report was classified until after the fall of the Soviet Union, and has been discovered in the former Soviet archives only recently. Procurement ratios are reported for each province and year from 1928 to 1932 in RSAE 4372/30/871 (p. 25). Dividing procurement levels by procurement ratios recovers production levels.

The main advantages of our correction are that it can be calculated at province levels and requires few, straightforward assumptions. Our correction assumes that a decline in procurement reflects a decline in production rather than a decline in the ability of the government to collect. If, for example, peasants hid more (less) grain to evade procurement in 1932 than in other years, then our estimates would understate (overstate) true production. Given historical accounts that intense campaigns to prevent peasants from hiding grain started with the introduction of collectivization several years before the famine, the *prima facie* assumption is that there was little evasion (and little change in evasion) between 1930 and 1932. In other analyses of the paper, we address the measurement error in production figures by not using reported production during this period.

The Gosplan report does not cover 1933. Thus, to estimate production for 1933, we follow the spirit of Davies and Wheatcroft (2009) and Tauger (1991) and calculate 1933 production as collectivization rate $\times$ grain sown area $\times$ kolkhoz yield + (1 - collectivization rate) $\times$ grain sown area $\times$ official yield. 1933 kolkhoz yields are reported in the Gosplan report on the state of collective farms in 1932–33 (RSAE 1562/77/70, p. 31). Collectivization rate is the share of collectivized rural households. This calculation assumes that sown area was similar for collectivized and non-collectivized households. In reality, collectivized households were given more land to farm. It also assumes that individual farm yields were equal to official yields.

Panel II rows (11) and (13)-(14) show that, except for 1930, our corrections fall within the DW range for years for which both are available. Given the conventional wisdom that official aggregate numbers were inflated only in the years when production fell relative to expectations, it is reassuring that our corrected production is below officially reported production levels in row (12) for 1931 to 1933.

[1] The derivation of these estimates, which are presented in Davies and Wheatcroft (2009) Table 1, p. 448-449, is unclear. The authors explain that "Our estimates are based on a range of different data that were accepted internally by the best experts of the time, and our own assessments of the reliability of these different data" (p. 446).

Note that after 1933, the government switched from reporting realized production to the maximum possible yields given weather, geographic conditions, sown area, labor and other inputs. This change did not systematically differ based on the share of ethnic Ukrainians in each region and should not affect our study.[2]

B Predicting Grain Production

To estimate the grain production function, we use data from 1901 to 1915. We regress log grain production (total harvests) on log province area, log FAO GAEZ grain suitability index, their interaction, temperature and precipitation for each of the four seasons, their pairwise interactions and squared terms (without a constant). The seasons are: fall (October, November, and December of the previous calendar year), winter (January, February, March), spring (April, May, June), and summer (July, August, September). The weather data are from Matsuura and Willmott (2014). Appendix Table A.4 presents the estimated grain production function. We then use this production function to predict grain harvest from 1922 to 1940. The predicted grain and actual grain are closely correlated, with two exceptions: Karelia and East Siberia provinces. The in-sample R-squared is 0.90. The out-of-sample R-squared is 0.77. The high out-of-sample predictive power is consistent with the lack of major technological changes in Soviet agriculture before the 1930s.

C Standard Errors

In Appendix Table A.5, we present spatially correlated standard errors with different cutoffs, and we allow for spatially correlated errors as well as those that are correlated over time and space. Our baseline follows the recommendations by Colella et al. (2019) in adjusting for spatial correlation within 1,500 kilometers (the mean province width in our sample is 1,300 km). The estimates are not very sensitive to these changes. We also present Newey-West standard errors, standard errors that are clustered at the year level, the province level, and the year and province levels. For these, we estimate wild bootstrapped standard errors to account for the small number of clusters (Cameron et al., 2008). The estimates are almost always statistically significant at the 10% or higher level. Note that in column (1), when we use the corrected grain estimates, we only have five years and there are too few clusters to estimate the wild bootstrapped standard errors that are clustered at the year level.

The wild bootstrap method is not appropriate for cases when there are a few number of clusters without the strong assumption of strict homogeneity between clusters (Canay et al., 2021). Note that the estimates using the district level panel can cluster the standard errors at the district level without wild bootstraps. The results are similar to the baseline when we estimate spatially correlated standard errors. In this case, we adjust for spatial correlation within 400 kilometers (the mean district width in our sample is 76 km). The province and district level results together give us confidence in our findings.

[2] Davies and Wheatcroft (2009) discusses this change from "the barn harvest" to "the harvest on the root."

D Robustness

D.1 Measurement

Weather Given the prominence of weather in studies of famine, we are cautious and control for it in several different ways in addition to the baseline control of weather-predicted production. Table A.7 Panel A directly controls for weather. Row (2) controls for the conditions emphasized by earlier studies such as Davies and Wheatcroft (2009) and Tauger (1991): spring and summer temperature and precipitation and the following year's winter temperature and precipitation. Row (3) controls for monthly temperature and precipitation and their squared terms. Row (4) controls for monthly temperature and precipitation and their interactions. Row (5) controls for monthly weather following the standard in the literature, where the weather shock indicator is equal to one if the month's temperature or precipitation is more than one standard deviation away from the long-term (1900–50) mean. Row (6) controls for the deviations from the long-term mean of monthly temperature and precipitation for the 12 months of year t. Finally, as in Rozenas and Zhukov (2019), row (7) controls for the monthly deviations of temperature and precipitation for years $t-1$ and t. These additional controls cause little change in the interaction coefficient of interest relative to the baseline in row (1).

Ukrainian Population Share and Mortality Rates The baseline examines total mortality rate because it is available for a longer time horizon than rural or urban mortality rates and we address the fact that most mortality was rural by controlling for urban population share and its interaction with the famine dummy variable. Table A.7 Panel B rows (8) and (9) show that the interaction estimate of interest is statistically zero when the dependent variable is urban mortality and similar to the baseline when it is rural mortality. They confirm that Ukrainian bias in famine mortality was mostly a rural phenomenon.

Next, we use the total population or urban population share of ethic Ukrainians on the right-hand-side instead of the baseline measure of rural ethnic Ukrainian share. The standardized coefficients presented in italics in rows (10) and (11) are similar. Rows (12) and (13) show that the results are also similar if we measured Ukrainian population share with the share of people whose mother tongue is Ukrainian, which we observe in the 1926 and 1897 censuses, or the 1897 measure of ethnicity.

D.2 Omitted Variables

The following exercises address concerns that pre-famine Ukrainian population share may be correlated with omitted variables that influence famine severity.

Demographic Structure To address the concern that higher mortality in ethnic Ukrainian areas may be due to differences in the demographic composition across regions (i.e., young children are more vulnerable), Appendix Table A.7 row (14) controls for the population gender ratio and the share of individuals aged ten and younger (as reported by the 1926 Population Census), each interacted with the famine indicator. Appendix Table A.7 Panel B rows (15) to (18) demonstrate robustness to many alternative controls for the age and gender.

Row (19) controls for 1928 mortality rates interacted with the famine indicator. This addresses potential mechanical mean reversion in mortality rates. Again, the ethnic Ukrainian interaction coefficient is nearly identical to the baseline.

Other Food Production, Grain Exports We follow existing studies of the Soviet famine and focus on the production of grain because it was the main agricultural commodity and source of calories for the population, and because there are no systematic disaggregated data for the production of other foods. If other food production was systematically different in Ukrainian areas, then our interpretation of the mortality-Ukrainian gradient would be confounded. To address this, we control for the suitability of each province for the cultivation of other crops. Since suitability is mostly determined by geography, we first address this by controlling for the interaction of latitude, longitude and the famine dummy, and the lower order interaction terms. Then, we control for the suitability for each of the other staple and cash crops — potato, wheat, rye, sugar beets, sunflowers, and flax — as predicted by the FAO GAEZ model interacted with the famine dummy. Appendix Table A.7 rows (20) and (21) present the estimates. The Ukrainian interaction coefficient is robust. To investigate whether access to fishing might have been worse in the Ukrainian areas, we construct a variable for the length of rivers and streams normalized by province size. Finally, it is possible that places located close to grain-exporting ports might be more vulnerable to grain requisitions and famine. To account for this, we use the list of all 1913 grain-exporting ports from Departament tamozhennykh sborov, ed (1914) and calculate the distance from each district centroid to the nearest port. For each province, we calculate the average distance across districts. Rows (22) and (23) show that the coefficient of interest is robust to these controls.

Social Capital and Other Slow-Moving, Historical Factors We also directly control for historical factors that could have affected famine severity. Appendix Table A.7 Panel D controls for the interactions of various proxies of pre-Soviet regional wealth: the value of agricultural equipment in 1910 and livestock in 1916 (Castañeda Dower and Markevich, 2018). It also controls for proxies of cultural norms, institutions and inequality: the share of serfs in 1858 (three years before the abolition of serfdom), the shares of Catholics and Orthodox Christians (the two major religion groups in Ukraine) from the 1897 Population Census, the share of peasant households in repartition communes (which required more cooperation amongst peasants) and the land Gini estimated from the 1905 Land Census, and per capita peasant revolts from 1895 to 1914. The main result is robust.

Alternative Measures of *Dekulakization* Table A.7 panel E shows that our main result is robust to different ways of controlling for the extent of *dekulakization* in the region. These measures are the number of exiled *kulak* households during 1930–31 (Davies and Wheatcroft, 2009, Table 28), the number of exiled *kulak* households during 1930–31 according to a secret police report in (Berelowitch and Danilov, eds, 2000-2012, Document 253), ex ante 1930 quotas for *kulak* exile, secret police estimates of total number of *kulaks* in the countryside, the number of arrested peasants, and their first principal component.

Local Differences in Administrative Capacity or Political Zealousness This section addresses the concern that Ukrainian population share may be correlated with local administrative capacity or political zealousness to enforce Bolshevik policies. Appendix Table A.7 row (39) controls for the share of votes for the Bolshevik Party in the 1917 Constituency Assembly election from Protasov et al. (2014). This was first and only election until the end of the Bolshevik rule and approximately 60% of the eligible voters turned out to vote. We interpret this as local support for Bolshevik policies.

Row (40) controls for the size of the Communist Party in each province just before the famine: the number of Communist Party Members (averaged over 1922, 1927 and 1931) per 1,000 individuals in each province. Party members were the key enforcers of state policy in the countryside and were responsible for grain procurement. Row (41) controls for the number of delegates who attended the 1930 Party Congress. The Congress was a showcase of support where all delegates voted in the affirmative for comprehensive collectivization. We interpret the number of communists and 1930 Congress delegates as proxies for the administrative capacity and political zealousness. All three variables are time invariant and we control for the interaction of each and the famine dummy variable. The Ukrainian interaction coefficient changes little with these additional controls. This goes against the concern that anti-Ukrainian bias in famine mortality is due to differences in local administrative capacity or political alignment with Moscow.

Random Permutation Another way to demonstrate that our results are unlikely to be driven by coincidence is to conduct random permutation tests. We alternately permute Ukrainian population shares and mortality rates across provinces (while preserving the mean and variance of the sample distribution) and estimate the same equation using randomly assigned measures. In each permutation, we find that the probability that our main result is due to coincidence is 1% or less. This is done for 10,000 iterations. See Appendix Figure A.2.

D.3 District Level

We do not observe age structure at the district level. Thus, we control for the gender ratio of the whole population to address potential differences in demographic structure. See Appendix Table A.8. We also conduct a random permutation exercise with the district-level data and alternately permute ethnic Ukrainian population share and 1933 mortality across districts. Like the province-level results, we find that our estimates are unlikely to be driven by coincidence. We estimate the specification as in Table 3 column (2) using the randomly assigned ethnic Ukrainian population share or 1933 mortality. This is done for 10,000 iterations. Comparing the distribution of the coefficients from the permutations to the baseline estimate (indicated by the vertical red line) in Appendix Figure A.2, we find that the probability that the latter is due to coincidence is less than 1%.

E Border Effects

We estimate the size of the decline in mortality at the Ukraine-Russia border by regressing excess mortality on a dummy variable that equals one if the district is in Russia. We do this with samples that include districts

within 100 to 250 kilometers on either side of the border and find that famine mortality rates were on average 25 to 36 per 1,000 lower in Russia. See Appendix Table A.9 Panel A.

Table A.1: Ethnic Composition in the USSR

	1926 Census						1939 Census	
	Total		Urban		Rural		Total	
	mln.	%	mln.	%	mln.	%	mln.	%
	A. All USSR							
Russians	77.8	53.1	16.6	63.5	61.2	50.8	99.6	58.4
Ukrainians	31.2	21.3	3.3	12.6	27.9	23.2	28.1	16.5
Belorussians	4.7	3.2	0.5	1.9	4.2	3.5	5.3	3.1
	B. USSR (Regression Sample, Subset of Panel A)							
Russians	77.1	57.2	16.1	67.9	61.1	54.9	94.8	65.0
Ukrainians	31.1	23.1	3.2	13.7	27.9	25.1	27.1	18.6
Belorussians	4.7	3.5	0.5	2.0	4.2	3.8	5.2	3.6
	C. "Grain-producing" Provinces (Subset of Panel B)							
Ukrainians	28.5	43.8	3.0	29.2	25.5	46.6	25.3	37.1
Russians	27.3	41.9	5.2	50.0	22.0	40.3	32.8	48.1
Tatars	2.2	3.4	0.2	2.0	2.0	3.6	2.9	4.3

Notes: These data are reported by the 1926 and 1939 Population Censuses. Panel C includes Bashkir ASSR, Central Black-Earth region, Crimea, Lower Volga, Middle Volga, North Caucasus, Tatar ASSR, and Ukraine.

Table A.2: Grain Availability (Russia and USSR)

	1927	1928	1929	1930	1931	1932	1933	1937	1939
	I. Russia								
(1) Total population (mln)	100.9	103.3	106.0	108.3	111.0	106.1	101.3	109.1	112.4
(2) Rural population (mln)	76.6	85.7	87.3	88.3	88.4	87.1	86.8	73.4	74.6
(3) Production (mln tons)	.	60.5	48.2	54.8	43.0	44.3	45.6	.	.
(4) Procurement (mln tons)	6.2	9.6	10.2	13.9	14.8	13.7	15.3	.	.
(5) Rural retention (kg/person/day)	.	1.628	1.192	1.269	0.873	0.961	0.956	.	.
(6) Rural retention if no procurement	.	1.935	1.511	1.699	1.333	1.393	1.438	.	.
(7) Rural and urban retention, no procurement (kg/person/day)	.	1.605	1.245	1.385	1.063	1.143	1.232	.	.
(8) Food needs for heavy labor (kg/person/day)	0.78	0.78	0.78	0.78	0.78	0.78	0.79	0.75	0.75
	II. USSR								
(9) Total population (mln)	147.0	150.5	154.2	157.5	160.5	163.3	165.8	162.0	165.5
(10) Rural population (mln)	120.7	124.3	126.6	128.0	128.3	127.0	127.1	110.1	110.6
(11) Production (mln tons)	.	73.4	71.8	83.3	64.5	57.1	67.3	.	.
(12) Official	74.1	73.3	71.7	83.5	69.5	69.9	89.8	120.3	100.9
(13) DW Min	.	.	.	73.0	57.0	55.0	70.0	.	.
(14) DW Max	.	.	.	77.0	65.0	60.0	77.0	.	.
(15) Procurement (mln tons)	11.1	10.8	16.1	22.2	22.8	19.0	23.7	31.9	30.9
(16) Rural retention (kg/person/day)	.	1.379	1.205	1.306	0.890	0.822	0.941	.	.
(17) Rural retention if no procurement	.	1.617	1.553	1.782	1.378	1.232	1.452	.	.
(18) Rural and urban retention, no procurement (kg/person/day)	.	1.336	1.275	1.449	1.101	0.959	1.113	.	.
(19) Food needs for heavy labor (kg/person/day)	0.78	0.78	0.78	0.78	0.78	0.77	0.77	0.76	0.76

Notes: Data for population, production and procurement are official statistics. Rows (3) and (11) report production revised by the authors using archival data (see text). Row (12) reports official production figures. Rows (13)-(14) report production revised by Davis and Wheatcroft (2004) (see their Table 1). Retention is the difference between our revised production and procurement. Food needs are calculated by the authors and take into account the demographic composition (e.g., age, gender, rural/urban) as reported in the population censuses. They are based on official guidelines for maximum caloric needs for each group as reported by Lositskij (1926, 1928). Panel I includes Russia. Panel II includes all USSR.

Table A.3: Correlates of Ukrainian Population Share

		A. Province Characteristic Summaries				B. Regress Each Province Characteristic on the Rural Share of Ethnic Ukrainians			
	Province Characteristic	Min	Mean	Sd	Max	Coef.	Std. Err.	Observations	R-squared
(1)	Official Grain Production 1928, kg/person/day	0.461	1.351	0.575	2.369	0.299	(0.415)	19	0.013
(2)	Potato suitability	0.051	0.476	0.209	0.690	0.157	(0.103)	19	0.026
(3)	Sugarbeet suitability	0.001	0.408	0.239	0.669	0.188	(0.119)	19	0.029
(4)	Sunflower suitability	0.000	0.411	0.257	0.829	0.423	(0.118)	19	0.126
(5)	Wheat suitability	0.048	0.523	0.266	0.864	0.378	(0.130)	19	0.094
(6)	Rye suitability	0.048	0.508	0.258	0.859	0.456	(0.126)	19	0.145
(7)	Flax suitability	0.003	0.501	0.292	0.855	0.405	(0.136)	19	0.089
(8)	Length of Rivers and Streams per Area, km/km-squared	0.058	0.084	0.014	0.114	0.015	(0.012)	19	0.052
(9)	Urbanization 1932	0.104	0.251	0.125	0.509	-0.016	(0.060)	19	0.001
(10)	Gender ratio (male/female) 1926	0.827	0.908	0.044	1.022	0.052	(0.020)	19	0.064
(11)	Share of childern under 10, 1926	0.245	0.286	0.020	0.324	-0.013	(0.011)	19	0.020
(12)	Nominal income pc 1897	0.042	0.075	0.032	0.171	0.004	(0.017)	19	0.001
(13)	Real income pc 1897	0.050	0.076	0.024	0.134	0.014	(0.015)	19	0.016
(14)	Labor Productivity 1897	0.100	0.147	0.042	0.228	0.037	(0.027)	19	0.037
(15)	Rural Labor Productivity 1897 (Lower Bound)	0.052	0.087	0.034	0.182	0.046	(0.033)	19	0.088
(16)	Rural Labor Productivity 1897 (Upper Bound)	0.154	0.263	0.095	0.532	0.120	(0.095)	19	0.074
(17)	Value of Agricultural Equipment 1910 (rub/person)	5.232	7.597	1.994	12.509	3.628	(3.039)	19	0.154
(18)	Livestock 1916 (heads/person)	0.880	1.613	0.835	3.726	0.020	(0.662)	19	0.00003
(19)	Share of Catholics 1897	0.001	0.013	0.023	0.099	0.014	(0.015)	19	0.017
(20)	Share of Orthodox Christians 1897	0.496	0.853	0.125	0.988	0.016	(0.056)	19	0.001
(21)	Share of Serfs 1858	0.000	0.250	0.206	0.603	0.049	(0.178)	19	0.003
(22)	Peasant Revolts 1895-1914 (per 1,000)	0.016	0.062	0.039	0.154	0.006	(0.022)	19	0.001
(23)	Share of Peasant Land in Repartition Commune 1905	0.433	0.918	0.175	0.999	-0.352	(0.169)	16	0.222
(24)	Share of Peasant Households in Repartition Commune 1905	0.451	0.912	0.177	1.000	-0.419	(0.190)	16	0.305
(25)	Peasant and Private Land Gini 1905	0.147	0.481	0.173	0.847	0.204	(0.134)	16	0.076
(26)	1917 Bolshevik Vote Share	0.036	0.218	0.155	0.540	-0.209	(0.082)	19	0.085
(27)	#Communists (averaged over 1922, 1927, and 1931) per 1,000	4.107	8.921	4.108	18.741	-1.621	(2.510)	19	0.007
(28)	#1930 Congress Delegates per 100,000	0.000	0.159	0.084	0.288	0.034	(0.054)	19	0.008

Notes: The sample includes Ukraine, Russia and Belarus. Observations are at the province level. Potato, sugarbeet, sunflower, wheat, rye and flax suitabilities are FAO GAEZ agricultural suitabilities for rain-fed low-input agriculture. Panel A: each row presents a summary of the province characteristic. Panel B: each row presents a separate regression of the province characteristic on the rural share of ethnic Ukrainians in the province. Huber-White robust standard errors are in parentheses.

Table A.4: The Effect of Weather and Natural Conditions on Grain Production

	Dependent Variable: Log Grain Production		
	(1)		(2)
Log area	0.352	Fall temperature × Fall precipitation	0.0005
	(0.067)		(0.0002)
Log grain suitability	-4.643	Winter temperature × Winter precipitation	0.001
	(0.640)		(0.0003)
Log area × Log grain suitability	0.278	Spring temperature × Spring precipitation	0.0004
	(0.023)		(0.0003)
Fall temperature	0.015	Summer temperature × Summer precipitation	0.001
	(0.037)		(0.0003)
Winter temperature	0.027	Fall temperature^2	0.004
	(0.043)		(0.002)
Spring temperature	-0.169	Winter temperature^2	0.0003
	(0.079)		(0.002)
Summer temperature	-0.978	Spring temperature^2	-0.001
	(0.194)		(0.003)
Fall precipitation	-0.006	Summer temperature^2	0.028
	(0.006)		(0.005)
Winter precipitation	-0.005	Fall precipitation^2	0.00002
	(0.007)		(0.00002)
Spring precipitation	0.01	Winter precipitation^2	0.00004
	(0.007)		(0.00003)
Summer precipitation	-0.025	Spring precipitation^2	-0.00003
	(0.009)		(0.00002)
		Summer precipitation^2	0.00003
			(0.00001)
Observations		220	
R-squared		0.998	

Notes: The sample includes Ukraine, Russia and Belarus. Observations are at the province and year level and include years from 1901 to 1914. Log grain is the logarithm of the grain harvest. Log area is the logarithm of province area. Log grain suitability is the logarithm of the province's FAO GAEZ wheat suitability index for rain-fed low-input agriculture. Fall is October, November, December of the previous calendar year; Winter is January, February, March; Spring is April, May, June; Summer is July, August, September.

Table A.5: Standard Errors of the Baseline Estimate of Ukrainians × Famine

	Dependent Variable: Mortality in Year $t+1$				
	Control for Corrected Grain Production	Baseline	Control for Kulaks × Famine	Control for Livestock Change × Famine	Omit Ukraine
	(1)	(2)	(3)	(4)	(5)
Ukrainians × Famine	0.050	0.051	0.054	0.053	0.086
Standard Errors:					
Spatial corr within 1500 km (baseline)	**(0.002)**	**(0.006)**	**(0.005)**	**(0.005)**	**(0.007)**
Spatial corr within 900 km	(0.003)	(0.007)	(0.006)	(0.006)	(0.007)
Spatial corr within 1,000 km	(0.003)	(0.006)	(0.006)	(0.006)	(0.007)
Spatial corr within 1,100 km	(0.003)	(0.006)	(0.006)	(0.006)	(0.007)
Spatial corr within 1,200 km	(0.003)	(0.006)	(0.006)	(0.006)	(0.008)
Spatial corr within 1,300 km	(0.002)	(0.006)	(0.006)	(0.006)	(0.008)
Spatial corr within 1,400 km	(0.002)	(0.006)	(0.006)	(0.006)	(0.008)
Spatial corr within 1,600 km	(0.002)	(0.006)	(0.005)	(0.005)	(0.007)
Spatial corr within 1,700 km	(0.002)	(0.006)	(0.005)	(0.005)	(0.007)
Spatial corr within 1,800 km	(0.002)	(0.006)	(0.005)	(0.005)	(0.007)
Spatial corr within 1,900 km	(0.002)	(0.006)	(0.005)	(0.005)	(0.007)
Spatial corr within 2,000 km	(0.002)	(0.006)	(0.005)	(0.005)	(0.007)
Spatial corr within 900 km and over time corr within 19 years	(0.003)	(0.007)	(0.006)	(0.006)	(0.008)
Spatial corr within 1,000 km and over time corr within 19 years	(0.003)	(0.007)	(0.006)	(0.006)	(0.008)
Spatial corr within 1,100 km and over time corr within 19 years	(0.003)	(0.006)	(0.006)	(0.006)	(0.008)
Spatial corr within 1,200 km and over time corr within 19 years	(0.003)	(0.006)	(0.006)	(0.006)	(0.008)
Spatial corr within 1,300 km and over time corr within 19 years	(0.003)	(0.006)	(0.006)	(0.006)	(0.008)
Spatial corr within 1,400 km and over time corr within 19 years	(0.003)	(0.006)	(0.006)	(0.006)	(0.008)
Spatial corr within 1,500 km and over time corr within 19 years	(0.003)	(0.006)	(0.006)	(0.006)	(0.008)
Spatial corr within 1,600 km and over time corr within 19 years	(0.003)	(0.006)	(0.006)	(0.006)	(0.008)
Spatial corr within 1,700 km and over time corr within 19 years	(0.003)	(0.006)	(0.005)	(0.005)	(0.007)
Spatial corr within 1,800 km and over time corr within 19 years	(0.003)	(0.006)	(0.005)	(0.005)	(0.007)
Spatial corr within 1,900 km and over time corr within 19 years	(0.003)	(0.006)	(0.005)	(0.005)	(0.007)
Spatial corr within 2,000 km and over time corr within 19 years	(0.003)	(0.006)	(0.005)	(0.005)	(0.007)
Huber-White robust standard errors	(0.004)	(0.008)	(0.007)	(0.007)	(0.008)
Bootstrapped p-value cluster(year)		0.0004	0.0004	0.001	0.003
Bootstrapped p-value cluster(province)	0.004	0.134	0.038	0.018	0.003
Bootstrapped p-value cluster(province year)	0.062	0.130	0.034	0.021	0.019
Observations	107	337	337	337	319
R-squared	0.838	0.776	0.791	0.784	0.758

Notes: The table reports standard errors and bootstrapped p-values of the Ukrainians × Famine coefficient for different levels of spatial and over time correlation; see Colella et al. (2019) for details.

Table A.6: The Dynamic Relationship between Ukrainian Population Share, Mortality and Grain Retention

	Dependent Variable:	
	Mortality in Year $t+1$ (1)	Retention (2)
Ukrainians × 1923	-0.002	
	(0.003)	
Ukrainians × 1924	-0.003	
	(0.003)	
Ukrainians × 1925	0.001	
	(0.002)	
Ukrainians × 1926	-0.001	
	(0.003)	
Ukrainians × 1927	-0.002	
	(0.004)	
Ukrainians × 1928	-0.001	
	(0.003)	
Ukrainians × 1929	0.002	-0.126
	(0.002)	(0.292)
Ukrainians × 1930	-0.002	0.114
	(0.004)	(0.247)
Ukrainians × 1931	0.007	-0.117
	(0.002)	(0.254)
Ukrainians × 1932	0.049	-1.174
	(0.007)	(0.393)
Ukrainians × 1933	-0.003	-0.455
	(0.003)	(0.358)
Ukrainians × 1934	-0.004	
	(0.003)	
Ukrainians × 1935	-0.010	
	(0.005)	
Ukrainians × 1936	-0.002	
	(0.003)	
Ukrainians × 1937	-0.003	
	(0.003)	
Ukrainians × 1938	-0.004	
	(0.003)	
Ukrainians × 1939	-0.004	
	(0.004)	
Observations	337	107
R-squared	0.811	0.801
Dep. var. mean	*0.022*	*1.000*

Notes: The sample includes Ukraine, Russia and Belarus. Observations are at the province and year level; column (1) includes years from 1922 to 1939, column (2) includes years from 1928 to 1933. Mortality is the number of deaths divided by the total population. Retention is the authors' revised grain production (see text) minus grain procurement, measured in kilograms per person per day. Ukrainians is the 1926 rural Ukrainian population share. All estimates control for urbanization interacted with year FE, and province and year FE. Column (1) also controls for grain (per capita grain production predicted by exogenous factors) interacted with year FE. The standard errors in parentheses are adjusted for spatial correlation within 1,500 km.

Table A.7: Famine Severity and Ethnic Ukrainian Population Share— Robustness to Alternative Measures and Additional Controls

Dependent Variable: Mortality in Year $t+1$						
	Dep. var. mean.	Ukrainians × Famine			Observations	R-squared
		Coef.	Std. Err.	Std. Coef.		
	(1)	(2)	(3)	(4)	(5)	(6)
(1) Baseline	0.022	0.051	(0.006)	*0.826*	337	0.776
A. Control for Weather instead of Predicted Grain						
(2) Spring, Summer, Winter Temp and Rain	0.022	0.051	(0.004)	*0.830*	337	0.782
(3) Monthly Temp, Rain, and Quadratics (48 vars)	0.022	0.051	(0.003)	*0.829*	337	0.828
(4) Monthly Temp, Rain, Temp × Rain (36 vars)	0.022	0.048	(0.003)	*0.785*	337	0.814
(5) Monthly Weather Shock (12 vars)	0.022	0.050	(0.004)	*0.818*	337	0.784
(6) Monthly Temp and Rain Deviations from Historical Monthly Mean in Year t (24 vars)	0.022	0.048	(0.003)	*0.781*	337	0.806
(7) Monthly Temp and Rain Deviations for Year t-1, t (48 vars)	0.022	0.047	(0.002)	*0.772*	337	0.825
B. Alternative Measures of Mortality, Ukrainian Share						
(8) Urban Mortality	0.021	0.013	(0.008)	*0.252*	285	0.746
(9) Rural Mortality	0.022	0.063	(0.005)	*0.869*	285	0.787
(10) Total Ukrainians	0.022	0.055	(0.006)	*0.820*	337	0.776
(11) Urban Ukrainians	0.022	0.090	(0.010)	*0.785*	337	0.771
(12) Mother Tongue Ukrainian 1926	0.022	0.056	(0.008)	*0.747*	337	0.765
(13) Mother Tongue Ukrainian 1897	0.022	0.058	(0.007)	*0.820*	337	0.781
C. Control for Demographic Structure, Suitability for Other Crops						
(14) Share of Infants × Famine, Gender Ratio × Famine	0.022	0.055	(0.006)	*0.901*	337	0.780
(15) Share of Children 5 and Younger × Famine, Gender Ratio × Famine	0.022	0.046	(0.007)	*0.756*	337	0.785
(16) Share of Children 10 and Younger × Famine, Gender Ratio × Famine	0.022	0.048	(0.007)	*0.781*	337	0.783
(17) Share of Adults 50 and Older × Famine, Gender Ratio × Famine	0.022	0.051	(0.005)	*0.833*	337	0.784
(18) Share of Adults 70 and Older × Famine, Gender Ratio × Famine	0.022	0.051	(0.005)	*0.837*	337	0.784
(19) Mortality 1928 x Famine	0.022	0.050	(0.006)	*0.819*	337	0.777
(20) Latitude × Longitude × Famine	0.022	0.059	(0.004)	*0.965*	337	0.806
(21) Famine × Suitability for: Potato, Sugarbeet, Sunflower, Wheat, Rye, Flax	0.022	0.056	(0.004)	*0.913*	337	0.806
(22) Length of Rivers and Streams per Area × Famine	0.022	0.049	(0.005)	*0.791*	337	0.778
(23) Log(Avg Distance to the Nearest Grain Exporting Port) × Famine	0.022	0.051	(0.005)	*0.824*	337	0.776
D. Control for Slow-Moving Economic, Cultural and Institutional Features						
(24) Value of Agricultural Equipment 1910 × Famine	0.022	0.046	(0.004)	*0.747*	337	0.810
(25) Livestock 1916 × Famine	0.022	0.051	(0.005)	*0.831*	337	0.777
(26) Share of Catholics 1897 × Famine, Share of Orthodox Christians 1897 × Famine	0.022	0.053	(0.006)	*0.869*	337	0.782
(27) Share of Serfs 1858 × Famine	0.022	0.052	(0.005)	*0.850*	337	0.779
(28) Peasant Revolts 1895-1914 × Famine	0.022	0.050	(0.006)	*0.819*	337	0.797
(29) Baseline with Info on Land 1905	0.021	0.042	(0.005)	*0.690*	286	0.792
(30) Share of Peasant Land in Repartition Commune 1905 × Famine	0.022	0.053	(0.006)	*0.854*	286	0.799
(31) Share of Peasant Households in Repartition Commune 1905 × Famine	0.022	0.056	(0.008)	*0.917*	286	0.800
(32) Peasant and Private Land Gini 1905 × Famine	0.022	0.040	(0.006)	*0.645*	286	0.799
E. Alternative Controls for *Dekulakization*						
(33) Exiled Kulaks (DW) × Famine	0.022	0.054	(0.005)	*0.876*	337	0.792
(34) Exiled Kulaks (OGPU) × Famine	0.022	0.053	(0.006)	*0.862*	337	0.789
(35) Planned Kulaks 1930 × Famine	0.022	0.044	(0.007)	*0.712*	337	0.807
(36) Total Kulaks (OGPU Estimate) × Famine	0.022	0.058	(0.007)	*0.944*	267	0.780
(37) Arrested Kulaks 1930 × Famine	0.022	0.051	(0.006)	*0.836*	302	0.787
(38) First Principal Component of Kulak Variables × Famine	0.022	0.053	(0.005)	*0.868*	249	0.800
F. Control for Administrative Capacity and Political Zealousness						
(39) 1917 Bolshevik Vote Share × Famine	0.022	0.050	(0.006)	*0.807*	337	0.781
(40) #Communists (averaged over 1922, 1927, and 1931) × Famine	0.022	0.051	(0.005)	*0.825*	337	0.776
(41) #1930 Congress Delegates × Famine	0.022	0.050	(0.006)	*0.815*	337	0.784

Notes: The sample includes Ukraine, Russia and Belarus. Observations are at the province and year level. Each row is one regression. All regressions control for urbanization, urbanization × famine, and province and year FE. Row (1) and Panels B-F also control for grain (per capita grain production predicted by exogenous factors) and grain × famine. In row (5), a shock is a dummy variable which equals 1 if temp or rain is one std. dev. or more different from historical province mean. The standard errors in parentheses are adjusted for spatial correlation within 1,500 km.

Table A.8: Famine Mortality and Ethnic Ukrainian Population Share — Robustness Checks for District-level Estimates

	Dependent Variable: Mortality									
		Alternative Measures of Mortality		Alternative Measures of Ukrainian Population Share			Additional Controls			
	Baseline with Province-Year FE	Urban Mortality	Rural Mortality	Total Ukrainians	Urban Ukrainians	Mother Tongue Ukrainian	Gender Ratio	Latitude × Longitude × Famine	Famine × Suitability for: Potato, Sugarbeet, Sunflower, Rye, Flax	Log(Distance to the Nearest Grain Exporting Port) × Famine
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ukrainians × Famine	0.024	0.0004	0.036	0.031	0.027	0.035	0.024	0.023	0.025	0.023
	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)
	[0.007]	[0.014]	[0.011]	[0.007]	[0.012]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]
Standardized Coef.	*0.311*	*0.007*	*0.441*	*0.384*	*0.276*	*0.444*	*0.311*	*0.294*	*0.324*	*0.303*
Observations	3,274	1,648	2,832	3,274	1,960	3,264	3,274	3,274	3,274	3,274
R-squared	0.812	0.913	0.920	0.814	0.840	0.816	0.813	0.814	0.818	0.812
Ukrainians										
Mean	*0.255*	*0.255*	*0.255*	*0.243*	*0.228*	*0.236*	*0.255*	*0.255*	*0.255*	*0.255*
Std. Dev.	*0.374*	*0.374*	*0.374*	*0.358*	*0.294*	*0.368*	*0.374*	*0.374*	*0.374*	*0.374*
Dep. var. mean in 1928	*0.019*	*0.017*	*0.020*	*0.019*	*0.019*	*0.019*	*0.019*	*0.019*	*0.019*	*0.019*
Dep. var. mean in 1933	*0.039*	*0.032*	*0.039*	*0.039*	*0.039*	*0.039*	*0.039*	*0.039*	*0.039*	*0.039*

Notes: The sample includes Ukraine and Russia. Observations are at the district and year level; the sample includes two years – 1928 and 1933. In columns (1) and (4)-(10), mortality is the number of deaths divided by the total population; in column (2), mortality is the number of urban deaths divided by the urban population; in column (3), mortality is the number of rural deaths divided by the rural population. In columns (1)-(3) and (7)-(10), Ukrainians is the 1926 rural Ukrainian population share; in column (4), Ukrainians is the 1926 total Ukrainian population share; in column (5), Ukrainians is the 1926 urban Ukrainian population share; in column (6), Ukrainians is the 1926 rural share of people whose mother tongue is Ukrainian. Famine is an indicator that equals to one in 1933 and zero otherwise. All regressions control for grain suitability × famine, monthly temperature and precipitation in years t and $t-1$ (48 additional controls), district and province × year fixed effects. Column (7) also controls for the 1926 male/female ratio × famine. Column (8) also controls for latitude × longitude × famine and all lower-order interactions. Column (9) also controls for agricultural suitability of each of the 5 crops listed in the column heading interacted with the famine dummy. Column (10) also controls for the log distance to the nearest grain exporting port × Famine. Grain suitability is the FAO GAEZ wheat suitability index for low-input rain-fed agriculture. The standard errors in parentheses are adjusted for spatial correlation within 400 km. Standard errors clustered at the district level are presented in square brackets.

Table A.9: Famine Mortality at the Ukrainian-Russian Border

Window, km	[-100,100] (1)	[-150,150] (2)	[-200,200] (3)	[-250,250] (4)
	A. Dependent Variable: Excess Mortality 1933			
Russia dummy	-0.025	-0.032	-0.036	-0.035
	(0.004)	(0.004)	(0.005)	(0.006)
Observations	144	211	274	336
R-squared	0.174	0.258	0.292	0.304
	B. Excess Mortality 1933 Demeaned by 1926 Ukrainian Share			
Russia dummy	0.001	-0.003	-0.003	-0.002
	(0.006)	(0.005)	(0.005)	(0.004)
Observations	144	211	274	336
R-squared	0.0004	0.004	0.005	0.002

Notes: The sample includes Russia and Ukraine. Observations are districts within the distance from Ukrainian-Russian border stated in the column headings. Excess mortality 1933 is the difference between 1933 and 1928 mortality for each district. In Panel B, excess mortality 1933 is demeaned by the 1926 rural share of ethnic Ukrainians in each district. The standard errors in parentheses are adjusted for spatial correlation within 400 km.

Table A.10: Grain Retention and Mortality

	Dependent Variable: Mortality in Year $t+1$		
	(1)	(2)	(3)
Retention: < 0.5	0.037	0.036	0.036
	(0.009)	(0.009)	(0.009)
Retention: (0.5, 1.0]	0.026	0.026	0.025
	(0.005)	(0.005)	(0.005)
Retention: (1.0, 1.5]	0.021	0.021	0.021
	(0.004)	(0.004)	(0.004)
Retention: (1.5, 2.0]	0.020	0.020	0.021
	(0.004)	(0.004)	(0.004)
Retention: > 2.0	0.016	0.015	0.015
	(0.005)	(0.005)	(0.005)
Controls:			
Urbanization $\times$ Famine		Y	Y
Official Grain 1928 $\times$ Famine			Y
Observations	107	107	107
R-squared	0.963	0.963	0.967
Dep. var. mean	*0.023*	*0.023*	*0.023*

Notes: The sample includes Ukraine, Russia and Belarus. Observations are at the province and year level for the years from 1928 to 1933. Mortality is the number of deaths divided by the total population. Retention is the authors' revised grain production (see text) minus grain procurement, measured in kilograms per person per day. All regressions control for urbanization, province and year fixed effects. Huber-White robust standard errors are in parentheses.

Table A.11: Population, Ethnic Composition, Agricultural Output and Mechanization after the Famine

	Dependent Variable:				
	A. Long-Term Population			B. Agriculture	
	Log Rural Population	Rural Share Ukrainians	Rural Share Russians	Log(Grain)	Tractors (Tractor Horse Power/1928 Sown Area)
	(1)	(2)	(3)	(4)	(5)
Ukrainians × 1897	-0.113	-0.058	0.044		
	(0.099)	(0.009)	(0.016)		
Ukrainians × 1922				-0.252	
				(0.242)	
Ukrainians × 1923				0.269	
				(0.210)	
Ukrainians × 1924				0.105	
				(0.208)	
Ukrainians × 1925				-0.189	
				(0.207)	
Ukrainians × 1926				-0.724	
				(0.123)	
Ukrainians × 1927				-0.353	
				(0.192)	
Ukrainians × 1928					0.001
					(0.0002)
Ukrainians × 1929				-0.333	0.003
				(0.180)	(0.001)
Ukrainians × 1930				-0.398	0.005
				(0.205)	(0.001)
Ukrainians × 1931				-0.144	0.007
				(0.201)	(0.002)
Ukrainians × 1932				-1.720	0.008
				(0.213)	(0.003)
Ukrainians × 1933				-0.772	0.016
				(0.264)	(0.005)
Ukrainians × 1934				-1.011	0.029
				(0.143)	(0.008)
Ukrainians × 1935				-0.783	0.029
				(0.169)	(0.008)
Ukrainians × 1936				0.113	0.039
				(0.325)	(0.012)
Ukrainians × 1937	-0.309			-0.364	0.024
	(0.071)			(0.158)	(0.007)
Ukrainians × 1938				-0.189	0.018
				(0.230)	(0.007)
Ukrainians × 1939	-0.220	-0.217	0.219	0.065	0.013
	(0.039)	(0.150)	(0.150)	(0.224)	(0.006)
Ukrainians × 1940				-0.222	
				(0.165)	
Ukrainians × 1959	0.013	-0.179	0.255		
	(0.093)	(0.155)	(0.177)		
Ukrainians × 1970	0.177	-0.184	0.270		
	(0.162)	(0.154)	(0.176)		
Ukrainians × 1979	0.312	-0.193	0.281		
	(0.207)	(0.151)	(0.173)		
Ukrainians × 1989	0.356	-0.214	0.284		
	(0.260)	(0.148)	(0.160)		
Ukrainians × 2002	0.701				
	(0.266)				
Observations	171	133	133	354	247
R-squared	0.952	0.941	0.957	0.985	0.899
Dep. var. mean	*14.754*	*0.086*	*0.663*	*9.939*	*0.027*

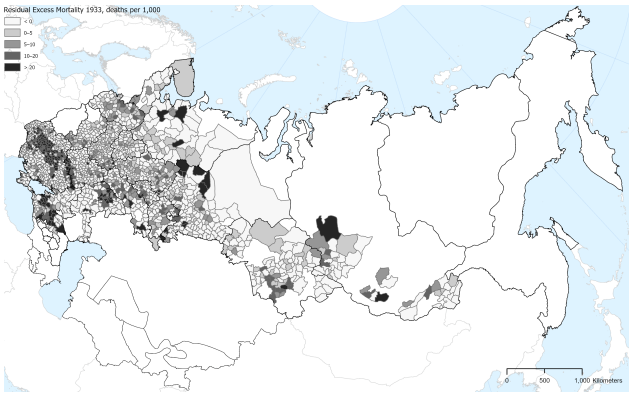
Notes: The sample includes Ukraine, Russia and Belarus. Observations are at the province and year level. Panel A includes all census years: 1897, 1926, 1937 (rural population only, no ethnic composition), 1939, 1959, 1970, 1979, 1989, and 2002 (rural population only, no ethnic composition). Panel B column (4) includes years from 1922 to 1940. Panel B column (5) includes years from 1927 to 1939. Ukrainians is the 1926 rural Ukrainian population share. Log(grain) is the officially reported grain production for years from 1922 to 1927 and from 1934 to 1940, and revised grain production (see text) for years from 1928 to 1933. All regressions control for province and year fixed effects. Column (4) also controls for monthly temperature and precipitation in year t and $t-1$ (48 additional controls). The standard errors in parentheses are adjusted for spatial correlation within 1,500 km.

Figure A.1: Maps

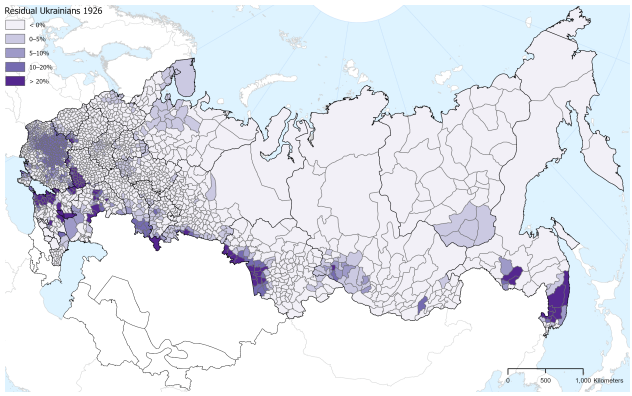
(a) Province Excess Mortality 1933 and Grain-Producing Regions

(b) Province Ethnic Ukrainians (1926) and Grain-Producing Regions

(c) District Excess Mortality 1933 Demeaned by Province Fixed Effects



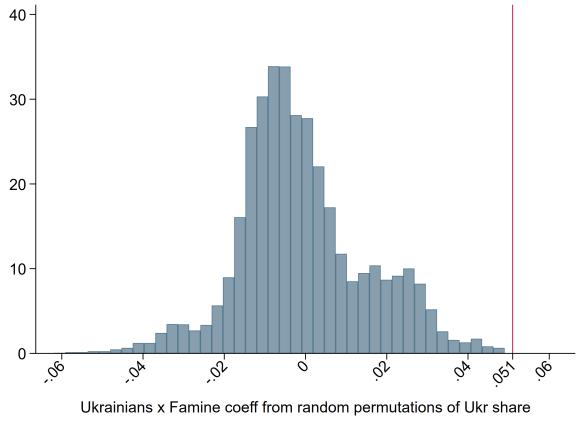
(d) District Ethnic Ukrainians 1926 Demeaned by Province Fixed Effects



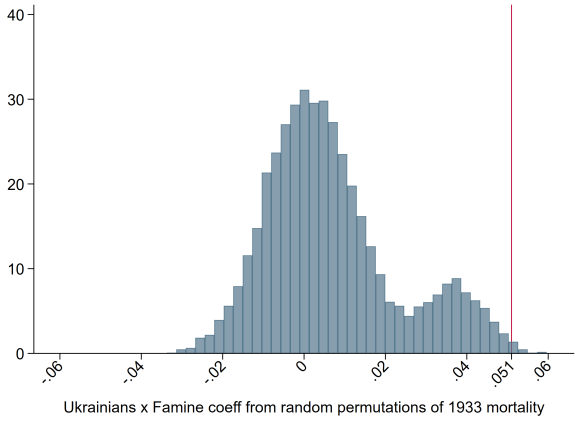
Notes: Excess mortality 1933 is mortality in 1933 minus mortality in 1928. Ethnic Ukrainians 1926 is the share of ethnic Ukrainians in the rural population according to the 1926 Population Census.

Figure A.2: Random Permutation Test

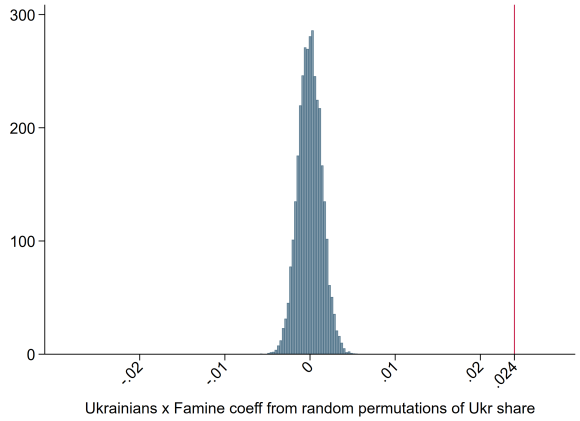
(a) Permute Ukrainian Population Share, Province-Level Panel



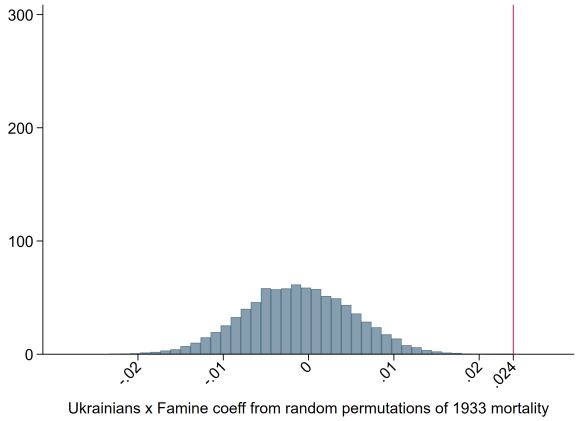
(b) Permute 1933 Mortality, Province-Level Panel



(c) Permute Ukrainian Population Share, District-Level Panel

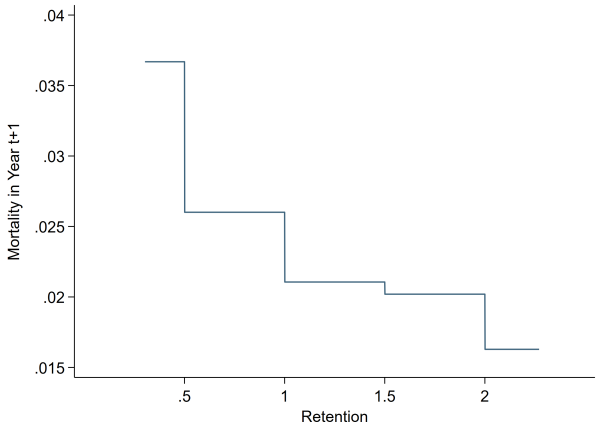


(d) Permute 1933 Mortality, District-Level Panel



Notes: The figures are histograms of the interaction coefficients ($Ukrainians_i \times Famine_t$) estimated from 10,000 iterations. The red vertical lines indicate the baseline estimates reported in the tables.

Figure A.3: Piece-wise Linear Estimation of Grain Retention and Mortality

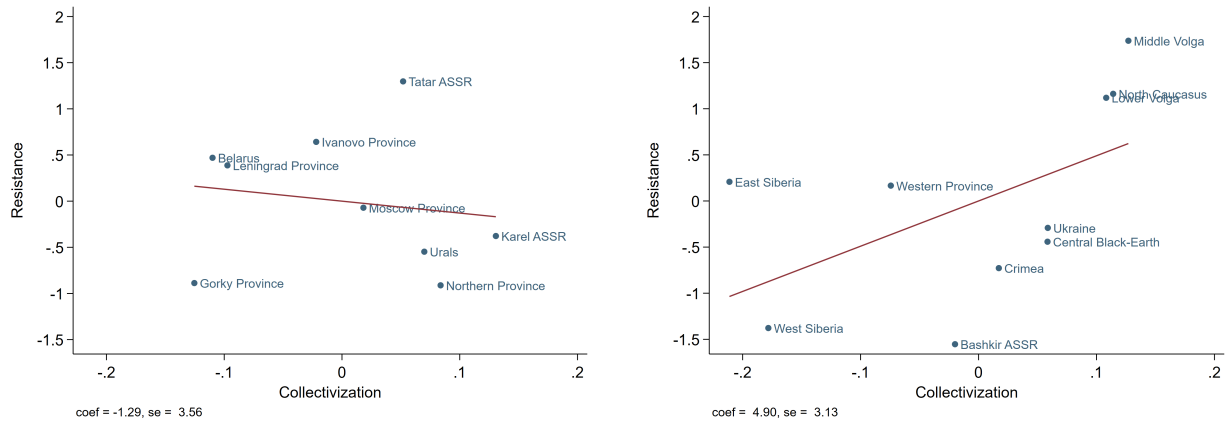


Notes: The figure plots the coefficients for different levels of grain retention (production minus procurement), controlling for urbanization and province and year fixed effects, from the piece-wise linear function in equation (2). The coefficients and standard errors are shown in Appendix Table A.10.

Figure A.4: Peasant Resistance to Collectivization

(a) Low Ukrainian Share

(b) High Ukrainian Share



Notes: Figures A.4a and A.4b plot the fitted lines and residuals for the correlation between collectivization and peasant resistance, controlling for official 1928 per capita grain production and urbanization. Figure A.4a includes provinces with Ukrainian share below the sample median. Figure A.4b includes provinces with Ukrainian share above or equal to the sample median. The p-value for the statistical difference of the two slopes is 0.093.

Data Appendix

We combine multiple archival and published sources to build the province and district-level datasets used in this paper. We provide the comprehensive list of sources below. The archival data that we manually collected or digitized for the first time are indicated with "**" in the heading.

Province-level Panel

The province-level panel includes 1922 to 1940 and 19 provinces within the republics of Belarus, Russia and Ukraine. These provinces correspond to the 1932 administrative division. Belarus and Ukraine were each a single province. The omitted territories are those with no reliable mortality data: Far Eastern Province, Yakut Autonomous SSR, and the North Caucasus ethnic territories: Chechen Autonomous Province, Cherkess Autonomous Province, Dagestan Autonomous SSR., Ingush Autonomous Province, Kabardino-Balkarian Autonomous Province, Karachay Autonomous Province, North Ossetian Autonomous Province. Figure A.1a maps the provinces in our sample. Omitted territories are in white.

Total, Rural and Urban Population For the years 1927, 1937 and 1939, we use data from population censuses.[3] For earlier years, 1920, 1923, 1924 and 1925 we use population estimates constructed by Soviet statisticians (see the exact sources below). For 1933, we use the same source as the 1933 district mortality: in addition to reporting the number of live births and deaths it also reports the population for the beginning of the year that was served by the civil acts registration bureaus (ZAGSy). For the remaining years, we interpolate population. Our main finding is not an artifact of the interpolation: the results are similar with only 1926 population census data (Markevich et al., 2023, Table 2, Column 9).

Administrative boundaries differ across years. However, since the data are reported at disaggregated administrative units, we are able to harmonize the data over time by using ArcGIS and manually aggregate the population data to the 1932 province borders. We use hand-created ArcGIS 1932 map. One issue is that small changes in borders that occur over time lead to large changes in population if we assume that the population is uniformly distributed across space in very large and sparsely populated provinces such as Ural, and West and East Siberia. To address this, we use the 1897 Population Census (the most recent available census prior to the start of our sample), which can be disaggregated to the *Uezd* level (of which there are 817 for the Russian Empire). These data allow us to calculate population density, which we use to attribute population to the 1932 province borders.

The data sources are as follows. 1920: Tsentralnoye Statisticheskoye Upravleniye [Central Statistical Office] (1926) *"Statisticheskiy yezhegodnik 1924 god (Vypusk pervyy) [Statistical Yearbook 1924 (First Issue)]"*, Volume VIII, Issue 7 of Trudy Tsentralnogo Statisticheskogo Upravleniya [Proceedings of the Central Statistical Office], Part I, Table 1.B. 1921-1922: total population is interpolated between 1920 and 1923; urban population is interpolated between 1920 and 1925. 1923: total population is calculated using the total number of deaths and deaths per 10,000 from Tsentralnoye Statisticheskoye Upravleniye [Central Statistical Office] (1926) *"Statisticheskiy yezhegodnik 1924 god (Vypusk pervyy) [Statistical Yearbook 1924 (First Issue)]"*, Volume VIII, Issue 7 of Trudy Tsentralnogo Statisticheskogo Upravleniya [Proceedings of the Central Statistical Office], Part I, Table 5; urban population is interpolated between 1920 and 1925. 1924: total population is calculated using the total number of deaths and deaths per 10,000 from Tsentralnoye Statisticheskoye Upravleniye [Central Statistical Office] (1926) *"Statisticheskiy yezhegodnik 1924 god (Vypusk pervyy) [Statistical Yearbook 1924 (First Issue)]"*, Volume VIII, Issue 7 of Trudy Tsentralnogo

[3]The 1926 Census took place on December 17, 1926, all other Soviet censuses took place in January; we use 1926 Census counts for the 1927 population. Population in the 1939 Census is widely believed to be inflated. This was recently corrected by Russian demographers using archival data (Bogoyavlensky, 2013). We use the corrected count.

Statisticheskogo Upravleniya [Proceedings of the Central Statistical Office], Part I, Table 8; urban population is interpolated between 1920 and 1925. 1925: Tsentralnoye Statisticheskoye Upravleniye [Central Statistical Office] (1926) *"Statisticheskiy yezhegodnik 1924 god (Vypusk pervyy) [Statistical Yearbook 1924 (First Issue)]"*, Volume VIII, Issue 7 of Trudy Tsentralnogo Statisticheskogo Upravleniya [Proceedings of the Central Statistical Office], Part I, Table 1.B. 1926: is interpolated between 1925 and 1927. 1927: December 17, 1926 Population Census from Demoscope.ru. 1928–1932: is interpolated between 1927 and 1933. 1933: Russian state archive of economy (hereafter, RGAE) 1562/329/19 p. 1–12. 1934–1936: is interpolated between 1933 and 1937. 1937: the 1937 Population Census from Zhiromskaya, V.B. and Kiselev, I.N. and Polyakov, Yu.A. (1996) *"Polveka pod grifom "sekretno": Vsesoyuznaya perepis naseleniya 1937 goda [Classified for half a century: All-Union population census of 1937]"*, Moscow: Nauka. 1938: is interpolated between 1937 and 1939. 1939: the 1939 Population Census corrected for the centralized additions (*pripiski*) from Demoscope.ru. 1940: used 1939 value. With the exception of the 1926 and 1939 population censuses, we digitized all sources mentioned in this paragraph for for the first time.

For the long-run estimates on population, we use multiple censuses taken from Demoscope.ru. For 1939, we use revised estimates for rural population provided by Russian demographers (Bogoyavlensky, 2013) because the official figures have been inflated. The revised estimates are not available by ethnicity. Each observation is a province in a census year. The data are harmonized to use 1932 province borders. Population size in the census reflects mortality, fertility and migration.

Natality and Mortality** Mortality is the total number of deaths divided by population (crude death rate). Natality is the number of live births divided by population (crude birth rate). We assign these data into 1932 province boundaries following the same procedure as for population. The rural-urban decomposition of deaths and births is available since 1926. We are the first to digitize all sources mentioned in this paragraph for 1923-1926.

The data sources are as follows. 1923: Tsentralnoye Statisticheskoye Upravleniye [Central Statistical Office] (1926) *"Statisticheskiy yezhegodnik 1924 god (Vypusk pervyy) [Statistical Yearbook 1924 (First Issue)]"*, Volume VIII, Issue 7 of Trudy Tsentralnogo Statisticheskogo Upravleniya [Proceedings of the Central Statistical Office], Part I, Table 5. 1924: Tsentralnoye Statisticheskoye Upravleniye [Central Statistical Office] (1926) *"Statisticheskiy yezhegodnik 1924 god (Vypusk pervyy) [Statistical Yearbook 1924 (First Issue)]"*, Volume VIII, Issue 7 of Trudy Tsentralnogo Statisticheskogo Upravleniya [Proceedings of the Central Statistical Office], Part I, Table 8. 1925: Tsentralnoye Statisticheskoye Upravleniye S.S.S.R. [Central Statistical Office of the USSR] (1928) *"Yestestvennoye dvizheniye naseleniya Soyuza S.S.R. 1923–1925 [Natural movement of the population of the USSR]"*, Volume I, Issue 1, Table 1. 1926: Yestestvennoye dvizheniye naseleniya Soyuza S.S.R. v 1926 g, Izdaniye TsSU SS.S.R. (1929), Table 1. 1927–1932: Belarus, Ukraine – RGAE 1562/329/256; Russia – Demoscope.ru. 1933–1940: Demoscope.ru.

Ethnic Composition Ethnic composition comes from the 1897 and the 1926 Population Censuses. The 1897 Census reports population by mother tongue. We use the share of people whose mother tongue is Belorussian, Russian (*Velikorusskiy*), and Ukrainian (*Malorusskiy*). The 1926 Census reports population by self-proclaimed ethnicity and by mother tongue, we use both. Data are calculated in our province borders using 1897 and hand-created district (*volost*)-level 1926 maps. The original disaggregated figures for the 1897 Census are taken from Demoscope.ru. The 1926 figures are digitized by the authors from published census volumes TsSU (1928-1930). The 1897 map is from Kessler, Gijs and Andrei Markevich, Electronic Repository of Russian Historical Statistics, 18th - 21st centuries, https://ristat.org/, Version I (2020). We hand-created the ArcGIS 1926 map using paper-maps from the 1926 census volumes.

Age Structure Region (*okrug*)-level population by one-year age groups from the 1926 Population Census is reported by Demoscope.ru. We calculated the share of people aged 10 and younger using hand-created region (*okrug*)-level map. This procedure is legitimate because regions (*okruga*) are smaller than our provinces.

Gender Ratio Male to female ratio is from the 1926 Population Census as reported by Demoscope.ru. We calculated it in our province borders using hand-created district (*volost*)-level 1926 map. This procedure is legitimate because districts (*volosty*) are smaller than our provinces.

Grain Harvest, Sown Area, and Yield** The data sources are as follows. 1901–1914: Obukhov V.M. (1927) *"Dvizheniye urozhayev zernovykh kultur v Yevropeyskoy Rossii v period 1883–1915 g.g. [Movement of grain crops in European Russia in the period 1883–1915]"* and *Yezhegodnik Rossii* 1904–1916. 1922: Tsentralnoye Statisticheskoye Upravleniye [Central Statistical Office] (1924) *"Sbornik statisticheskikh svedeniy po Soyuzu S.S.R. 1918–1923. Za pyat let raboty Tsentralnogo Statisticheskogo Upravleniya [A collection of statistical information on the USSR 1918–1923. Five years of work of the Central Statistical Office.]"*, Volume XVIII of Trudy Tsentralnogo Statisticheskogo Upravleniya [Proceedings of the Central Statistical Office], Part VI, Tables 7 and 8. 1923: Tsentralnoye Statisticheskoye Upravleniye [Central Statistical Office] (1924) *"Statisticheskiy yezhegodnik 1922 i 1923 g. (Vypusk pervyy) [Statistical Yearbook 1922 and 1923 (First Issue)]"*, Volume VIII, Issue 5 of Trudy Tsentralnogo Statisticheskogo Upravleniya [Proceedings of the Central Statistical Office], Part III, Tables 3 and 4. 1924: Tsentralnoye Statisticheskoye Upravleniye [Central Statistical Office] (1926) *"Statisticheskiy yezhegodnik 1924 god (Vypusk pervyy) [Statistical Yearbook 1924 (First Issue)]"*, Volume VIII, Issue 7 of Trudy Tsentralnogo Statisticheskogo Upravleniya [Proceedings of the Central Statistical Office], Part III, Tables 6 and 7. 1925–1927: Statisticheskoye izdatelstvo TsSU SS.S.R. [Statistical Publishing House of the Central Statistical Office of the USSR] (1929) *"Selskoye khozyaystvo SS.S.R. 1925–1928. Sbornik statisticheskikh svedeniy k XVI Vsesoyuznoy partkonferentsii [Agriculture of the USSR 1925–1928. A collection of statistical information for the XVI All-Union Party Congress]"*, Part III. 1928: RGAE 1562/329/1409. 1929–1930: Gosudarstvennoye sotsialno-ekonomicheskoye izdatelstvo [State Socio-Economic Publishing House] (1932) *"Narodnoye khozyaystvo SS.S.R.. Statisticheskiy spravochnik 1932 [The national economy of the USSR. Statistical Handbook 1932]"*, Part II.3.A, Tables 30 and 33. 1931: Gosudarstvennoye izdatelstvo kolkhoznoy i sovkhoznoy literatury "Selkhozgiz" [State publishing house of collective and state farm literature "Selkhozgiz"] (1936) *"Selskoye khozyaystvo SS.S.R.. Yezhegodnik 1935 [Agriculture of the USSR. Yearbook 1935]"*, p. 269, Tables 106 and 107. 1932–1940: RGAE 1562/329/1409. The years 1922, and 1924 to 1927 are reported for larger units than our provinces. We map the grain data into 1932 provinces borders following the same procedure as for population.

Procurement** The data sources are as follows. 1924: Tsentralnoye Konventsionnoye Byuro Khlebozagotoviteley [Central Conventional Bureau of Grain Procurers] (1928) *"Yezhegodnik khlebnoy torgovli N1 [Yearbook of grain trade N 1]"*, Table 6. 1925: Tsentralnoye Konventsionnoye Byuro Khlebozagotoviteley [Central Conventional Bureau of Grain Procurers] (1928) *"Yezhegodnik khlebnoy torgovli N1 [Yearbook of grain trade N 1]"*, Table 14. 1926: Tsentralnoye Konventsionnoye Byuro Khlebozagotoviteley [Central Conventional Bureau of Grain Procurers] (1928) *"Yezhegodnik khlebnoy torgovli N1 [Yearbook of grain trade N 1]"*, Table 22. 1927: Statisticheskoye izdatelstvo TsSU SS.S.R. [Statistical Publishing House of the Central Statistical Office of the USSR] (1929) *"Selskoye khozyaystvo SS.S.R. 1925–1928. Sbornik statisticheskikh svedeniy k XVI Vsesoyuznoy partkonferentsii [Agriculture of the USSR 1925–1928. A collection of statistical information for the XVI All-Union Party Congress]", Part V.* 1928: calculated from the 1928 grain harvest and procurement as a share of harvest from RGAE 4372/30/871 p. 30. 1929: Narodnyy Komissariat

Snabzheniya SS.S.R. [People's Commissariat of Supply of the USSR] (1932) *"Yezhegodnik khlebooborota N4 [Yearbook of grain turnover N 4]"*, Tables 3 and 10. 1930: Narodnyy Komissariat Snabzheniya SS.S.R. [People's Commissariat of Supply of the USSR] (1932) *"Yezhegodnik khlebooborota N4 [Yearbook of grain turnover N 4]"*, Table 29 and Table 36. 1931: Komitet po zagotovkam S.-Kh produktov pri SNK SS.S.R. [Committee for Procurement of Agricultural Products under the Council of People's Commissars of the USSR] (1934) *"Yezhegodnik khlebooborota za 1931-32, 1932-33 i predvaritelnyye itogi zagotovok 1933 g. [Yearbook of grain turnover for 1931-32, 1932-33 and preliminary results of procurement in 1933]"*, Table 21. 1932: Komitet po zagotovkam S.-Kh produktov pri SNK SS.S.R. [Committee for Procurement of Agricultural Products under the Council of People's Commissars of the USSR] (1934) *"Yezhegodnik khlebooborota za 1931-32, 1932-33 i predvaritelnyye itogi zagotovok 1933 g. [Yearbook of grain turnover for 1931-32, 1932-33 and preliminary results of procurement in 1933]",* Table 33. 1933: Komitet po zagotovkam S.-Kh produktov pri SNK SS.S.R. [Committee for Procurement of Agricultural Products under the Council of People's Commissars of the USSR] (1934) *"Yezhegodnik khlebooborota za 1931-32, 1932-33 i predvaritelnyye itogi zagotovok 1933 g. [Yearbook of grain turnover for 1931-32, 1932-33 and preliminary results of procurement in 1933]",* Table 53. 1925 to 1927 data are more disaggregated than our provinces. Thus, we use hand-created ArcGIS maps to create the panel.

Grain Targets The First Five-Year plan reports plans for population, grain production, and procurement ("marketable grain"). Aggregate Soviet-level figures are available for every year from 1928 to 1933; from these yearly figures it is clear that the planners just linearly interpolated targets between 1928 and 1933. Disaggregated province-level figures are available for 1928 and 1933; we linearly interpolate planned grain production, procurement and population targets between these years. Summing up the interpolated province targets mechanically results in the aggregate target for that year. Conceptually, the interpolation assumes that the central planners used the same formula to set targets at the aggregate and province levels.

The provinces in the Plan are slightly larger than the provinces in our sample (e.g., West and East Siberia are reported together, Ivanovo and Moscow provinces are united into a Central Industrial Region, Leningrad province and Karelia are united, Tatar and Chuvas regions are united with the Middle Volga region). We calculate grain production and procurement targets in kilograms per person per day (planned grain divided by planned population). Planovoye khozyaystvo [Planned economy] (1930) *"Pyatiletniy plan narodno-khozyaystvennogo stroitel'stva SSSR. Tom 3: rayonnyy razrez plana [Five-year plan for the national economic construction of the USSR. Volume 3: regional aspect of the plan.]"*

Collectivization 1931: Izd. Kolkhoztsentra SS.S.R. i RSFSR [Publishing House of the Collective Farm Center of the USSR and the RSFSR] (1931) *"Kolkhoznoye stroitelstvo v SS.S.R. [Collective farms building in the USSR]"*, p. 15 and Davies and Wheatcroft (2009), Table 27. 1932: RGAE 1562/82/271. 1933: *"Plan. Zhurnal Gosplana i TsUNKhU SS.S.R. [Plan. Journal of the State Planning Committee and TsUNKhU USSR]"*, 2-1933.

Dekulakization The baseline measure of *kulak* households exiled during 1930 to 1931 per 1930 population is estimated as the average between Exiled *kulaks* (DW) and Exiled *kulaks* (OGPU) defined below. Exiled *kulaks* (DW) is the number of *dekulakized* and exiled households in Category II of *kulaks* in 1930 to 1931 according to Davies and Wheatcroft (2009) (Table 28) per 1930 population. Exiled *kulaks* (OGPU) is the number of *dekulakized* and exiled households of all categories between January 1, 1930 and July 1, 1931 according to an OGPU (secret police) 1931 report per 1930 population. The report is published in Berelovich A. and V. Danilov (2003). *"Sovetskaya derevnya glazami VChk-OGPU-NKVD. 1918—1939. Documents i materialy"* [Soviet Countryside from the Perspective of VChK-OGPU-NKVD]. Moscow: Rosspen. Vol. 3 *"*1930—1934 gg.*"*, Book 1. *"*1930—1931 gg.*"*, document 253.

Planned *kulaks* (lower bound) and Planned *kulaks* (upper bound) is the OGPU (secret police) planned number of *dekulakizations* by as of February, 1930 per 1930 population. The planned figures are published in Danilov, Victor, Robert Manning and Lynne Viola (Eds.). (1999-2006). *"Tragediya Sovetskoj Derevni. Kollektivizatsiya i raskulachivanie. Dokumenti i materialy v 5 tomakh, 1927-1939"* [Tragedy of the Soviet Countryside. Collectivization and Dekulakization. Documents and Materials. 5 volumes]. Moscow: Rosspen. Volume 2 "November 1929 — December 1930", Document 69. Total *kulaks* (OGPU estimate) is the total number of *kulaks* in the rural population according to the OGPU (secret police) estimate published in Berelovich A. and V. Danilov (2003). *"Sovetskaya derevnya glazami VChk-OGPU-NKVD. 1918—1939. Documents i materialy"* [Soviet Countryside from the Perspective of VChK-OGPU-NKVD]. Moscow: Rosspen. Vol. 3 "1930—1934 gg.", Book 1. "1930—1931 gg.", document 253. Arrested *kulaks* 1930 is the number of peasants processed by "troiki" in 1930 per 1930 population according to the OGPU (secret police) estimate published in Berelovich A. and V. Danilov (2003). *"Sovetskaya derevnya glazami VChk-OGPU-NKVD. 1918—1939. Documents i materialy"* [Soviet Countryside from the Perspective of VChK-OGPU-NKVD]. Moscow: Rosspen. Vol. 3 "1930—1934 gg.", Book 1. "1930—1931 gg.", document 279.

Peasant Resistance to the Soviet Regime** Information on peasant resistance is from previously classified secret police reports. They include three types of resistance. The first is the number of anti-Soviet "violent acts" per 1,000 people, which include murders or attempted murders of local officials, arsons and the destruction of collective farm or state property. The second is the number of mass demonstrations in the countryside. The third is the number of episodes when anti-Soviet leaflets were distributed in the countryside. The analysis examines the first principal component of the three indicators on a cross-section of provinces. The results are similar if we examine each measure separately.

The sources are the following."Violent acts", mass demonstrations, and anti-Soviet leaflets registered by the OGPU (secret police) between January 1, 1930 and April 1, 1932 per 1,000 1930 population are according to two OGPU reports. The reports are published in Berelovich A. and V. Danilov (2003). *"Sovetskaya derevnya glazami VChk-OGPU-NKVD. 1918—1939. Documents i materialy"* [Soviet Countryside from the Perspective of VChK-OGPU-NKVD]. Moscow: Rosspen. Vol. 3 "1930—1934 gg.", Book 1. "1930—1931 gg.", document 272, and Danilov, Victor, Robert Manning and Lynne Viola (Eds.). (1999-2006). *"Tragediya Sovetskoj Derevni. Kollektivizatsiya i raskulachivanie. Dokumenti i materialy v 5 tomakh, 1927-1939"* [Tragedy of the Soviet Countryside. Collectivization and Dekulakization. Documents and Materials. 5 volumes]. Moscow: Rosspen. Volume 3 "Late 1930 — 1933", Document 118.

Peasant Resistance to the Tsarist Regime Peasant revolts from 1895 to 1914 are from Gokmen and Kofanov (2020).

Bolshevik Votes 1917 Bolshevik vote share is from Castañeda Dower and Markevich (2023), and orignally from Protasov et al. (2014). Data is calculated in our province borders using district (*uezd*)-level 1917 map from Castañeda Dower and Markevich (2023).

Communists** The 1922 and 1927 data use different administrative divisions. We harmonized them using hand-created ArcGIS maps.

The data sources are as follows. 1922: Izdatelskoye otdeleniye TsK RKP [Publishing Department of the Central Committee of the RCP] (1922) *"Vserosssiyskaya perepis chlenov RKP 1922 goda [All-Russian census of the members of the RCP in 1922]"*, Issue 3, Table 6. 1927: Statisticheskiy otdel TsK VKP(b) [Statistical Department of the Central Committee of the CPSU(b)] (1927) *"Vsesoyuznaya partiynaya perepis 1927*

goda. Chislennyy sostav VKP(b) na 10 yanvarya 1927 g. [All-Union Party Census of 1927. The composition of the CPSU(b) on January 10, 1927]", Issue 1. 1931: Tsentralnyy Komitet VKP(b). Organizatsionno-instruktorskiy otdel [Central Committee of the CPSU(b). Organizational and instructor department] (1932) *"Sostav VKP(b) v tsifrakh. Dinamika osnovnykh pokazateley rosta parti za 1930 i pervoye polugodiye 1931 g. [Composition of the CPSU(b) in numbers. Dynamics of the main indicators of the growth of the party for 1930 and the first half of 1931]"*.

Voting Delegates 1930** We collected location and ethnicity of all 1930 Party Congress delegates that served as province-, district-, city-, or borough-level Party secretary from Rossiyskiy Gosudarstvennyy Arkhiv Sotsial'no-Politicheskoy Istorii (Russian State Archive of Socio-Political History, RGASPI), Fund 58, Register 1, Files 1–16.

Religious Composition Religious composition is from the 1897 Population Census, available at Kessler, Gijs and Andrei Markevich, Electronic Repository of Russian Historical Statistics, 18th - 21st centuries, https://ristat.org/, Version I (2020).

Shares of Repartition Commune and Private Land Data on commune and private land ownership are originally from the 1905 Land Census. We calculate province shares from district (*uezd*)-level figures taken from Castañeda Dower and Markevich (2018), using 1897 and 1932 ArcGIS maps. The 1897 map is from Kessler, Gijs and Andrei Markevich, Electronic Repository of Russian Historical Statistics, 18th - 21st centuries, https://ristat.org/, Version I (2020).

Pre-Soviet Wealth Measures We estimate the value of agricultural machines by multiplying the number of agricultural machines of different types by their prices and taking the sum. Agricultural machines data are originally from the 1910 Census of Agricultural Machines. We calculate province shares from district (*uezd*)-level figures taken from Castañeda Dower and Markevich (2018), using hand-created ArcGIS district (*uezd*)-level maps. Prices are from Ministerstvo Zemledeliya [Ministry of Agriculture] (1917). *"Sbornik statistiko-ekonomicheskikh svedenij po sel'skomu khozyajstvu Rossii i inostrannikh gosudarstv. [A collection of statistical and economic information about agriculture in Russian and foerign countries]"*, Volume X. Horses, cows, and livestock in 1916 are originally from the 1916 Agricultural Census. We calculate province shares from district (*uezd*)-level figures taken from Castañeda Dower and Markevich (2018), using hand-created ArcGIS district (*uezd*)-level maps.

District-level Panel

District-level dataset includes district for which we were able to collect mortality data for 1933 in the republics of Russia and Ukraine. Appendix Figure A.1c maps the districts in our sample (omitted territories are in white). To visualize information at the district level, we hand-created ArcGIS district (*volost*)-level 1926 and 1934 maps. To construct the ArcGIS 1926 map, we use paper-maps from the 1926 census volumes (TsSU, 1928-1930). For the ArcGIS 1934 map, we photographed the maps from TsSUNKhU (1934). 1933 mortality was reported using January 1, 1934 administrative borders.

Mortality** We are digitized 1928 and 1933 mortality. 1928: Russia: State archive of the Russian federation (GARF) 374/23/7, 13, 31–32, 67, 72–91, 132, 158; Ukraine: Tsentralna Statistichna Uprava USRR [Central Statistical Office of Ukraine] (1929) "Ukraina: Statisticheskiy Schorichnik 1929 [Ukraine: Statistical Yearbook 1929]." 1933: RGAE 1562/329/18–19 (Wheatcroft and Garnaut, 2013).

Ethnic Composition** Ethnic composition comes from the 1926 Population Census. This census reports population by self-proclaimed ethnicity and by mother tongue, we digitized both at the most desegregated (*volost*)-level from published census volumes (TsSU, 1928-1930). Data is calculated in the 1934 district borders using hand-created ArcGIS district (*volost*)-level 1926 and 1934 maps.

Urbanization** 1928: used value from December 1926 Population Census. This census reports district (*volost*)-level rural population and, separately, the population of each urban settlement. To calculate rural and urban population in 1934 administrative borders, we use hand-created ArcGIS district (*volost*)-level 1926 and 1934 maps and located all urban settlements on the map. 1933: RGAE 1562/329/18–19. We digitized the 1933 urban and total population figures directly from the archival records. Note that Wheatcroft and Garnaut (2013) first discovered the 1933 district-level population data in the former Soviet archives.

Gender Ratio** Gender ratio is a ratio of males to females according to the 1926 Population Census. For that, we digitized the most disaggregated (*volost*)-level from published census volumes TsSU (1928-1930). To calculate data in 1934 administrative borders, we use hand-created ArcGIS district (*volost*)-level 1926 and 1934 maps.